(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 4 270 232 A1

(12) EUROPEAN PATENT APPLICATION
published in accordance with Art. 153(4) EPC

(43) Date of publication:
01.11.2023 Bulletin 2023/44

(21) Application number: **22773378.9**

(22) Date of filing: **11.08.2022**

(51) International Patent Classification (IPC):
G06F 21/62 *(2013.01)* ***G06T 3/00*** *(2006.01)*

(86) International application number:
PCT/CN2022/111704

(87) International publication number:
WO 2023/168903 (14.09.2023 Gazette 2023/37)

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME
Designated Validation States:
KH MA MD TN

(30) Priority: **10.03.2022 CN 202210234385**

(71) Applicant: **Tencent Technology (Shenzhen) Company Limited**
Shenzhen, Guangdong, 518057 (CN)

(72) Inventor: **The designation of the inventor has not yet been filed**

(74) Representative: **Gunzelmann, Rainer**
Wuesthoff & Wuesthoff
Patentanwälte PartG mbB
Schweigerstraße 2
81541 München (DE)

(54) MODEL TRAINING METHOD AND APPARATUS, IDENTITY ANONYMIZATION METHOD AND APPARATUS, DEVICE, STORAGE MEDIUM, AND PROGRAM PRODUCT

(57) This application provides a model training method and apparatus, an identity anonymization method and apparatus, a device, a storage medium, and a program product, which may be applied to various scenarios such as cloud technologies, artificial intelligence, intelligent traffic, and aided driving. The method includes: performing sampling on a target space of a projection module in a target network model to obtain N virtual identity vectors; performing, by using an attribute module in the target network model, attribute vector extraction on a to-be-processed image to obtain M attribute vectors; and performing, by using a fusion module in the target network model, image generation based on the N virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the to-be-processed image.

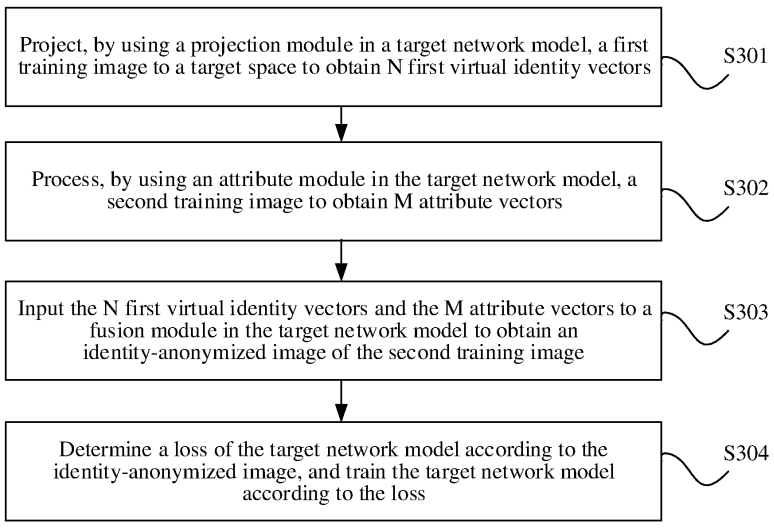

FIG. 3

EP 4 270 232 A1

Description

RELATED APPLICATION

[0001] This application claims priority to Chinese Patent Application No. 202210234385.2 filed on March 10, 2022, which is incorporated herein by reference in its entirety.

FIELD OF THE TECHNOLOGY

[0002] This application relates to the field of image processing technologies, and in particular, to a model training method and apparatus, an identity anonymization method and apparatus, a device, a storage medium, and a program product.

BACKGROUND OF THE DISCLOSURE

[0003] Identity anonymization is also referred to as de-identification, and is to remove a recognizable identity feature in an image or a video, but meanwhile keep other identity-free attributes unchanged, and ensure that an anonymized picture or video is still visually real.

[0004] In a related technology, a conditional generative adversarial network (GAN) is used for generating an anonymized picture, and a posture key point of an original picture is extracted, and the posture key point of the original picture and a background picture with a face area removed are inputted to a model as conditions, to generate a new virtual identity to fill the vacant face area. However, in this method, the background picture with the face area removed is used as model input, and therefore a picture generated by the model has poor quality.

SUMMARY

[0005] Embodiments of this application provide a model training method and apparatus, an identity anonymization method and apparatus, a computing device, a computer-readable storage medium, and a computer program product, to improve quality of a generated identity-anonymized image.

[0006] An embodiment of this application provides a model training method, including:

projecting, by using a projection module in a target network model, a first training image to a target space to obtain N first virtual identity vectors, N being a positive integer;

performing, by using an attribute module in the target network model, attribute vector extraction on a second training image to obtain M attribute vectors, M being a positive integer;

performing, by using a fusion module in the target network model, image generation based on the N first virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the second training image; and

determining a loss of the target network model according to the identity-anonymized image, and training the target network model according to the loss.

[0007] An embodiment of this application further provides an identity anonymization method, including:

performing sampling on a target space of a projection module in a target network model to obtain N virtual identity vectors, N being a positive integer;

performing, by using an attribute module in the target network model, attribute vector extraction on a to-be-processed image to obtain M attribute vectors, M being a positive integer; and

performing, by using a fusion module in the target network model, image generation based on the N virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the to-be-processed image.

[0008] An embodiment of this application further provides a model training apparatus, including:

a projection unit, configured to project, by using a projection module in a target network model, a first training image to a target space to obtain N first virtual identity vectors, N being a positive integer;

an attribute unit, configured to perform, by using an attribute module in the target network model, attribute vector extraction on a second training image to obtain M attribute vectors, M being a positive integer;

a fusion unit, configured to perform, by using a fusion module in the target network model, image generation based on the N first virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the second training image; and

a training unit, configured to determine a loss of the target network model according to the identity-anonymized image, and train the target network model according to the loss.

[0009] An embodiment of this application further provides an identity anonymization apparatus, including:

a sampling unit, configured to perform sampling on a target space of a projection module in a target network model to obtain N virtual identity vectors, N being a positive integer;

an attribute unit, configured to perform, by using an attribute module in the target network model, attribute vector extraction on a to-be-processed image to obtain M attribute vectors, M being a positive integer; and

an anonymization unit, configured to perform, by using a fusion module in the target network model, image generation based on the N virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the to-be-processed image.

[0010] An embodiment of this application further provides a computing device, including a processor and a memory. The memory is configured to store a computer program. The processor is configured to invoke and run the computer program stored in the memory to perform the model training method or the identity anonymization method provided in the embodiments of this application.

[0011] An embodiment of this application further provides a chip, configured to implement the model training method or the identity anonymization method provided in the embodiments of this application. The chip includes: a processor, configured to call and run a computer program from the memory, so that the device with the chip installed executes the foregoing model training method or identity anonymization method provided in the embodiments of this application.

[0012] An embodiment of this application further provides a computer-readable storage medium. The computer-readable storage medium is configured to store a computer program, the computer program, when executed, implementing the model training method or identity anonymization method provided in the embodiments of this application.

[0013] An embodiment of this application further provides a computer program product, including computer program instructions, when the computer program instructions are executed by a computer, the model training method or identity anonymization method in any one of the foregoing embodiments is implemented.

[0014] An embodiment of this application further provides a computer program, when run on a computer, performing the model training method or identity anonymization method according to the foregoing embodiments.

[0015] This embodiment of the present invention has the following beneficial effects:

[0016] In the training process of the target network model, the first training image is projected to the target space to obtain the N first virtual identity vectors, so that the target network model can fully learn identity information in the image; attribute vector extraction is performed on the second training image to obtain the M attribute vectors, so that the target network model fully learns attribute information in the image; and image generation is performed based on the N first virtual identity vectors and the M attribute vectors to obtain the identity-anonymized image of the second training image. In this way, a trained model can generate an image carrying virtual identity information while ensuring that attribute information of an original image remains unchanged;

[0017] In the application process of the target network model, sampling is performed on the target space of the projection module to obtain the N virtual identity vectors, so that virtual identity information is generated; attribute vector extraction is performed on the to-be-processed image to obtain the M attribute vectors, to ensure that an attribute feature in the to-be-processed image is not lost, and therefore ensure quality of the generated identity-anonymized image; and image generation is performed based on the N virtual identity vectors and the M attribute vectors to obtain the identity-anonymized image of the to-be-processed image, so that an identity-anonymized image carrying virtual identity information, that is, with a real identity hidden, is generated while ensuring that attribute information of the to-be-processed image remains unchanged. That is, in the embodiments of this application, during identity anonymization, an independent virtual identity is generated by using the target network model, without removing a face area in an image, thereby improving fidelity and a resolution of identity anonymization.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018]

FIG. 1A is a schematic diagram of a real image according to an embodiment of this application.

FIG. 1B to FIG. 1D are schematic diagrams of an identity-anonymized image corresponding to FIG. 1A according to an embodiment of this application.

FIG. 2 is a schematic diagram of a system architecture according to an embodiment of this application.

FIG. 3 is a schematic flowchart of a model training method according to an embodiment of this application.

FIG. 4 to FIG. 6 are schematic structural diagrams of a target network model according to an embodiment of this application.

FIG. 7 is a schematic structural diagram of a fusion module according to an embodiment of this application.

FIG. 8 is a schematic structural diagram of a target network model according to an embodiment of this application.

FIG. 9 and FIG. 10 are schematic diagrams of determining a contrast loss according to an embodiment of this application.

FIG. 11 is a schematic flowchart of an identity anonymization method according to an embodiment of this application.

FIG. 12 is a schematic diagram of a projection module according to an embodiment of this application.

FIG. 13 is a schematic diagram of determining an identity-anonymized image according to an embodiment of this application.

FIG. 14 is a schematic block diagram of a model training apparatus according to an embodiment of this application.

FIG. 15 is a schematic block diagram of an identity anonymization apparatus according to an embodiment of this application.

FIG. 16 is a schematic block diagram of a computing device according to an embodiment of this application.

DESCRIPTION OF EMBODIMENTS

[0019] The following clearly and completely describes the technical solutions in the embodiments of this application with reference to the accompanying drawings in the embodiments of this application.

[0020] It should be understood that in the embodiments of this application, "B corresponding to A" indicates that B is associated with A. In an implementation, B may be determined based on A. However, it is to be further understood that determining B according to A does not mean that determining B only according to A, and B may be determined according to A and/or other information.

[0021] In the descriptions of embodiments of this application, unless otherwise described, "a plurality of" means two or more than two.

[0022] In addition, for ease of describing the technical solutions in the embodiments of this application clearly, in the embodiments of this application, terms such as "first", "second", and "third" are used to distinguish same or similar items with a basically same function and purpose. A person skilled in the art may understand that the terms such as "first", "second", and "third" do not define a quantity and an execution sequence, and the terms such as "first", "second", and "third" do not indicate a definite difference.

[0023] For ease of understanding the embodiments of this application, related concepts involved in the embodiments of this application are first briefly introduced below.

[0024] Artificial intelligence (AI) is a theory, method, technology, and application system that uses a digital computer or a machine controlled by the digital computer to simulate, extend, and expand human intelligence, perceive an environment, obtain knowledge, and use knowledge to obtain an optimal result. In other words, AI is a comprehensive technology in computer science. This technology attempts to understand the essence of intelligence and produce a new

intelligent machine that can react in a manner similar to human intelligence. AI is to study the design principles and implementation methods of various intelligent machines, so that the machines can perceive, infer, and make decisions.

[0025] The AI technology is a comprehensive subject, relating to a wide range of fields, and involving both hardware and software techniques. Basic AI technologies generally include technologies such as sensors, dedicated AI chips, cloud computing, distributed storage, big data processing technologies, operating/interaction systems, and mechatronics. An AI software technology mainly includes fields such as a CV technology, a speech processing technology, a natural language processing technology, and machine learning/deep learning (DL).

[0026] ML is a multi-field interdiscipline, and relates to a plurality of disciplines such as the probability theory, statistics, the approximation theory, convex analysis, and the algorithm complexity theory. ML specializes in studying how a computer simulates or implements a human learning behavior to acquire new knowledge or skills, and reorganize an existing knowledge structure, so as to keep improving its performance. The ML is the core of the AI, is a basic way to make the computer intelligent, and is applied to various fields of AI. The ML and DL generally include technologies such as an artificial neural network, a belief network, reinforcement learning, transfer learning, inductive learning, and learning from demonstrations.

[0027] A method in the embodiments of this application may be applied to any scenario in which an image needs to be anonymized. For example, as shown in FIG. 1A to FIG. 1D, FIG. 1A shows a real image, and FIG. 1B to FIG. 1D show an identity-anonymized image of FIG. 1A. Through comparison between FIG. 1A and FIG. 1B to FIG. 1D, it can be learned that, in FIG. 1B to FIG. 1D, a recognizable identity feature in FIG. 1A is removed, while other identity-free attributes are kept unchanged, and it is ensured that the image is still visually real.

[0028] Scenario 1: The embodiments of this application may be applied to a privacy protection scenario. For example, for a face-related picture or video, the method in the embodiments of this application may be used for replacing a real identity with a virtual identity, so that subsequent tasks such as detection can be further performed without privacy leakage. In addition, when publishing a picture or a video, a user may also use the method in the embodiments of this application to an identity feature of the user, to avoid leakage of real information.

[0029] Scenario 2: The embodiments of this application may be applied to a scenario of generating a virtual image. For example, the technical solutions in the embodiments of this application may be used for generating a virtual identity, for example, a fixed latent identity variable, and replacing a background picture, to generate pictures or videos of a specific virtual image in different scenarios.

[0030] The scenario 1 and the scenario 2 are described by using an example in which a target is a face. The method in the embodiments of this application may be alternatively applied to a scenario of performing identity anonymization on another non-face target, for example, performing identity anonymization on any target such as an animal or a vehicle in a to-be-processed image.

[0031] In some embodiments, the method in the embodiments of the application may be applied to an intelligent transport system. An intelligent transport system (Intelligent Transport System, ITS) is also referred to as an intelligent transportation system (Intelligent Transportation System) and applies advanced technologies (an information technology, a computer technology, a data communications technology, a sensor technology, an electronic control technology, an automatic control technology, an artificial intelligence technology, and the like) to transportation, service control, and vehicle manufacturing comprehensively and effectively, so as to strengthen a connection between a vehicle, a road, and a user, thereby forming an integrated transportation system for safety assurance, efficiency improvement, environmental enhancement, and energy saving. For example, a solution combined with intelligent traffic in this application may be as follows: An in-vehicle device collects a face image of a user, performs identity anonymization on the collected face image by using the method in the embodiments of this application, and then transmits an identity-anonymized image to another device for performing task analysis or the like, for example, performing illegal driving analysis or intelligent driving analysis.

[0032] FIG. 2 is a schematic diagram of a system architecture according to an embodiment of this application. The system architecture includes user equipment 101, a data collection device 102, a training device 103, an execution device 104, a database 105, a content library 106, an I/O interface 107, and a target network model 108.

[0033] The data collection device 102 is configured to read training data from the content library 106, and store the read training data to the database 105. The training data involved in this embodiment of this application includes a first training image, a second training image, and a third training image, and the first training image, the second training image, and the third training image are all used for training the target network model.

[0034] In some embodiments, the user equipment 101 is configured to perform an annotation operation on data in the database 105.

[0035] The training device 103 trains the target network model 108 based on the training data maintained in the database 105, so that a trained target network model 108 can generate an identity-anonymized image of a to-be-processed image. In some embodiments, the target network model 108 obtained by the training device 103 may be applied to different systems or devices.

[0036] In FIG. 2, the execution device 104 is equipped with an I/O interface 107, for exchanging data with an external

device, for example, receiving, by using the I/O interface, a to-be-processed image transmitted by the user equipment 101. A computing module 109 in the execution device 104 processes the inputted to-be-processed image by using the trained target network model 108 to output an identity-anonymized image, and outputs the generated identity-anonymized image to the user equipment 101 for display, or inputs the generated identity-anonymized image to another task model for performing other task processing.

[0037] The user equipment 101 may include a mobile phone, a tablet computer, a notebook computer, a palmtop computer, a mobile Internet device (MID), or another terminal device with a function of installing a browser.

[0038] The execution device 104 may be a server. There may be one or more servers. When there are a plurality of servers, there may be at least one of the following cases: At least two servers are configured to provide different services; or at least two servers are configured to provide the same service, for example, provide the same service in a load balancing manner. This is not limited in this embodiment of this application. The server may be an independent physical server, or may be a server cluster or a distributed system formed by a plurality of physical servers, or may be a cloud server that provides basic cloud computing services such as a cloud service, a cloud database, cloud computing, a cloud function, cloud storage, a network service, cloud communication, a middleware service, a domain name service, a security service, a content delivery network (CDN), big data, and an AI platform. Alternatively, the server may become a node of a blockchain.

[0039] In this embodiment, the execution device 104 is connected to the user equipment 101 through a network. The network may be a wireless or wired network, for example, an intranet, the Internet, a Global System for Mobile Communications (GSM) network, a wideband code division multiple access (WCDMA) network, a 4G network, a 5G network, a Bluetooth network, a Wi-Fi network, or a call network.

[0040] FIG. 2 is only a schematic diagram of a system architecture according to an embodiment of this application, and a position relationship between the devices, the components, the modules, and the like shown in the figure does not constitute any limitation. In some embodiments, the data collection device 102, the user equipment 101, the training device 103, and the execution device 104 may be the same device. The database 105 may be distributed on one or more servers, and the content library 106 may be distributed on one or more servers.

[0041] Some embodiments are used below to describe in detail the technical solutions of the embodiments of this application. The following embodiments may be mutually combined, and same or similar concepts or processes may not be repeatedly described in some embodiments.

[0042] This application provides a target network model, and the target network model is used for performing identity anonymization on a target (for example, a face) in a to-be-processed image to generate an identity-anonymized image of the to-be-processed image. Therefore, in some embodiments, the target network model may be referred to as an identity anonymization model or an identity anonymizer.

[0043] First, a training process of the target network model is described.

[0044] FIG. 3 is a schematic flowchart of a model training method according to an embodiment of this application. In this embodiment of this application, the method is performed by an apparatus with a model training function, for example, a model training apparatus, and the model training apparatus may be a computing device, or a part of a computing device. The following provides descriptions by using an example in which the method is performed by a computing device. As shown in FIG. 3, the method of this embodiment of this application includes:

[0045] S301: The computing device projects, by using a projection module in a target network model, a first training image to a target space to obtain N first virtual identity vectors, N being a positive integer.

[0046] In this embodiment of this application, the first training image is a training image in training data. In a case that the first training image is a face image, the first training image is obtained with permission by a user, and collection, use, and processing of related image data need to comply with related laws, regulations, and standards of related countries and regions.

[0047] In this embodiment of this application, processes of training models by using first training images are basically similar. For ease of description, one first training image is used as an example for description.

[0048] In this embodiment of this application, the first training image is projected to the target space by using the target network model to obtain one or more virtual identity vectors of the first training image, so that the target network model learns identity information of the first training image. After the target network model fully learns the identity information, during actual identity anonymization, the target space of the target network model may be directly sampled to generate a virtual identity vector.

[0049] This embodiment of this application mainly relates to the following concepts: an attribute vector and a virtual identity vector.

[0050] The virtual identity vector is a vector corresponding to virtual identity information, and the virtual identity information is identity information with a recognizable identity feature hidden, for example, face information with a recognizable identity feature of a face hidden.

[0051] The attribute vector is a vector corresponding to attribute information, and feature information other than a recognizable identity feature in an image is referred to as attribute information, for example, background information.

[0052] In this embodiment of this application, the target network model may generate an independent virtual identity vector.

[0053] FIG. 4 is a schematic structural diagram of a target network model according to an embodiment of this application. As shown in FIG. 4, the target network model in this embodiment of this application includes a projection module, an attribute module, and a fusion module.

[0054] The projection module is configured to project a first training image to a target space to obtain N first virtual identity vectors of the first training image. N is a positive integer. A value of N is not limited in this embodiment of this application, and may be set according to an actual requirement.

[0055] The attribute unit is configured to perform attribute vector extraction on a second training image to obtain M attribute vectors of the second training image. M is a positive integer. A value of M is not limited in this embodiment of this application, and may be set according to an actual requirement. In some embodiments, M is equal to N.

[0056] The fusion module is configured to perform image generation based on the N first virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the second training image.

[0057] In a case that N is a positive integer greater than 1, the N first virtual identity vectors respectively correspond to different resolutions.

[0058] It can be learned from the foregoing descriptions that, in the target network model in this embodiment of this application, the projection module is configured to generate a virtual identity vector of a target in the second training image, a real identity feature of the target in the second training image being hidden in the virtual identity vector; and the attribute module is configured to extract an attribute vector of the second training image, features other than the real identity feature of the target in the second training image being retained in the attribute vector. In this way, after the fusion module performs image generation based on the virtual identity vector and the attribute vector, an anonymized image in which an identity of the target in the second training image is hidden, namely, an identity-anonymized image, can be obtained.

[0059] In some embodiments, as shown in FIG. 5, the projection module includes a first projection unit and a second projection unit, and the target space includes a first space Z and a second space W. In this case, the computing device may implement, in the following manner by using the projection module in the target network model, the projecting a first training image to a target space to obtain N first virtual identity vectors:

[0060] extracting priori identity information of the first training image; projecting, by using the first projection unit, the priori identity information to the first space Z to obtain N latent identity vectors; and projecting, by using the second projection unit, the N latent identity vectors to the second space W to obtain the N first virtual identity vectors.

[0061] As shown in FIG. 5, first, the priori identity information of the first training image is extracted, for example, the priori identity information of the first training image is extracted by using a pre-trained recognition model; then the priori identity information of the first training image is projected, by using the first projection unit, to the first space Z to obtain the N latent identity vectors; and then the N latent identity vectors is projected, by using the second projection unit, to the second space W to obtain the N first virtual identity vectors.

[0062] The first space Z and the second space W may be different latent spaces. The first space Z and the second space W are not limited in this embodiment of this application.

[0063] In some embodiments, the first space is a latent space Z, and the latent space Z conforms to a standard Gaussian distribution.

[0064] In this case, the first projection unit may project, in the following manner, the priori identity information to the first space Z to obtain the N latent identity vectors:

projecting, by using the first projection unit, the priori identity information as a mean and a variance of the first space; and performing sampling based on the mean and the variance of the first space to obtain the N latent identity vectors.

[0065] In some embodiments, the first projection unit is a variational autoencoder (VAE), for example, a conditional variational autoencoder (CVAE). The conditional variational autoencoder is a generative network, and learns a data distribution by using an encoder to obtain a latent variable, and then restores the latent variable to an original form of data by using a decoder. The conditional variational autoencoder may learn a data distribution and then perform sampling to generate new data, and is usually used for image generation.

[0066] In this way, the priori identity information of the first training image may be inputted to the VAE, and the VAE projects the priori identity information as a mean and a variance of the first space. Then sampling is performed based on the mean and the variance of the first space to obtain the N latent identity vectors of the first training image.

[0067] In this example, the first space is the latent space Z conforming to a standard Gaussian distribution. Therefore, to enhance an expression ability of the latent space, in this embodiment of this application, different latent vectors are generated at different resolution layers, for example, the N latent identity vectors are generated. This is equivalent to constructing a Z + space including a plurality of latent identity vectors.

[0068] In some embodiments, the second space W is obtained based on the latent space Z, for example, obtained by performing linear or nonlinear mapping on the latent space Z.

[0069] A network structure of the second projection unit is not limited in this embodiment of this application. For

example, a mapping network is used, and the mapping network includes a plurality of fully connected layers.

[0070] In this embodiment of this application, the priori identity information of the first training image is projected to a latent space (namely, the target space) of the projection module, so that the projection module fully learns identity information of the first training image, for subsequently generating a virtual identity vector that meets an actual requirement.

[0071] S302: Perform, by using an attribute module in the target network model, attribute vector extraction on a second training image to obtain M attribute vectors, M being a positive integer.

[0072] The second training image is any image in a training data set, and the second training image and the first training image may be the same image or different images.

[0073] In this embodiment of this application, the attribute module is configured to learn attribute information of the second training image to generate the M attribute vectors.

[0074] A network model of the attribute module is not limited in this embodiment of this application.

[0075] In some embodiments, as shown in FIG. 6, the attribute module includes an encoding unit and a decoding unit. In this case, attribute vector extraction may be performed on the second training image in the following manner by using the attribute module in target network model, to obtain the M attribute vectors:

inputting the second training image to the encoding unit to obtain feature information of the second training image; and inputting the feature information to the decoding unit to obtain the M attribute vectors.

[0076] In some embodiments, the encoding unit includes a plurality of feature extraction layers, the decoding unit also includes a plurality of feature extraction units, and a skip connection is implemented between at least one feature extraction layer in the encoding unit and at least one feature extraction layer in the decoding unit.

[0077] After the N first virtual identity vectors and the M attribute vectors are generated according to the foregoing steps, the following S303 is performed.

[0078] S303: Perform, by using a fusion module in the target network model, image generation based on the N first virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the second training image.

[0079] Example 1: The N first virtual identity vectors are spliced to obtain a spliced first virtual identity vector, the M attribute vectors are spliced to obtain a spliced attribute vector, and the spliced first virtual identity vector and the spliced attribute vector are used for performing image generation and then inputted to the fusion module to generate the identity-anonymized image.

[0080] For example, the spliced first virtual identity vector and the spliced attribute vector are concatenated and then inputted to the fusion module to generate the identity-anonymized image.

[0081] For another example, the spliced first virtual identity vector and the spliced attribute vector are added up and then inputted to the fusion module to generate the identity-anonymized image.

[0082] Example 2: The fusion module includes a plurality of different resolution layers. In this case, the fusion module may perform image generation in the following manner based on the N first virtual identity vectors and the M attribute vectors to obtain the identity-anonymized image of the second training image:

according to resolutions corresponding to the N first virtual identity vectors, inputting the N first virtual identity vectors to a corresponding resolution layer as patterns, and inputting the M attribute vectors to a corresponding resolution layer as noises, to obtain the identity-anonymized image of the second training image.

[0083] For example, N is 3, M is 4, and the fusion module includes four different resolution layers. The three first virtual identity vectors are denoted as a first virtual identity vector 1, a first virtual identity vector 2, and a first virtual identity vector 3. The four attribute vectors are denoted as an attribute vector 1, an attribute vector 2, an attribute vector 3, and an attribute vector 4. The four resolution layers are denoted as a resolution layer 1, a resolution layer 2, a resolution layer 3, and a resolution layer 4 in sequence according to magnitudes of resolutions. The first virtual identity vector 1 corresponds to the resolution layer 1 and the resolution layer 2 with a lower resolution, the first virtual identity vector 2 corresponds to the resolution layer 3 with a medium resolution, and the virtual identity vector 3 corresponds to the resolution layer 4 with the highest resolution. The four attribute vectors correspond to the four resolution layers in sequence according to magnitudes of resolutions.

[0084] For example, the first virtual identity vector 1 is inputted to the resolution layer 1 to obtain feature information 1; the attribute vector 1 and the feature information 1 are combined and then inputted to the resolution layer 2 along with the first virtual identity vector 1 to obtain feature information 2; the attribute vector 2 and the feature information 2 are combined and then inputted to the resolution layer 3 along with the first virtual identity vector 3 to obtain feature information 3; the attribute vector 3 and the feature information 3 are combined and then inputted to the resolution layer 4 along with the first virtual identity vector 4 to obtain feature information 4; and finally, the feature information 4 and the attribute vector 4 undergo processing such as combination to generate the identity-anonymized image of the second training image.

[0085] In some embodiments, the fusion module is a style-based generator (StyleGAN2). As shown in FIG. 7, an AdaIN layer is included between two adjacent resolution layers of the fusion module. For example, affine transform (AT) is performed on a first virtual identity vector i+1; feature information i outputted by the i^{th} resolution layer and an attribute

vector i are combined and then inputted to the AdaIN layer along with a first virtual identity vector i+1 obtained through affine transform; an AdaIN operation is performed, and an AdaIN operation result is inputted to the $(i+1)^{th}$ resolution layer.

[0086] The fusion module in this embodiment of this application may be alternatively an adversarial model, for example, a StyleGAN3 or a ProGAN. In a case that different adversarial models are used as the fusion module, the identity-anonymized image of the second training image may be determined in different manners. This is not limited in this embodiment of this application.

[0087] In some embodiments, the model training process in this embodiment of this application is described by using an example in which the first projection unit is a VAE, the second projection unit is a mapping network, the attribute module is an autoencoder, and the fusion module is a StyleGAN2.

[0088] For example, as shown in FIG. 8, priori identity information is generated based on a first training image Xs by using a pre-trained face recognition model. Then the priori identity information is inputted to the VAE, and the VAE projects the priori identity information to the first space Z to obtain N latent identity vectors, for example, obtain three N latent identity vectors, the three N latent identity vectors respectively corresponding to three different resolutions: low, medium, and high. Then the N latent identity vectors are inputted to the mapping network, and the mapping network projects the N latent identity vectors from the first space Z to the second space W to obtain N first virtual identity vectors. In addition, a second training image Xt is inputted to the autoencoder, and the autoencoder processes the second training image Xt to generate M attribute vectors. Finally, the M attribute vectors are inputted to layers of the StyleGAN2 as noises, and the N first virtual identity vectors are inputted to layers of the StyleGAN2 as patterns, to obtain an identity-anonymized image Ys,t, outputted by the StyleGAN2, of the second training image.

[0089] According to the foregoing steps, the first training image and the second training image are inputted to the target network model to obtain the identity-anonymized image, outputted by the target network model, of the second training image. Then the following S304 is performed to train the target network model.

[0090] S304: Determine a loss of the target network model according to the identity-anonymized image, and train the target network model according to the loss.

[0091] According to the foregoing steps, the target network model outputs the identity-anonymized image of the second training image, and the loss of the target network model is determined according to the identity-anonymized image.

[0092] In some embodiments, the identity-anonymized image is inputted to a judgment model, and the judgment model is a pre-trained model capable of predicting an anonymization degree of the identity-anonymized image. For example, the identity-anonymized image is inputted to the judgment model, and the judgment model performs identity recognition on the identity-anonymized image, and determines a recognition result as the loss of the target network model. In a case that recognition accuracy is high, an anonymization effect of the current target network model is not desired. In this case, a parameter in the target network model is adjusted according to the loss of the target network model. Then a new first training image and a new second training image are selected for performing the foregoing steps S301 to S304, to continue to train the target network model until the target network model meets a training end condition. The training end condition includes at least a quantity of training times reaching a preset quantity of times, or an anonymization degree of the model reaching an expected effect.

[0093] In some embodiments, in a case that the first space Z shown in FIG. 5 is a latent space conforming to a standard Gaussian distribution, in this embodiment of this application, a KL divergence constraint L_{kl} is applied to the N latent identity vectors in the first space Z to ensure that identity information is projected to a standard Gaussian distribution.

[0094] Based on this, in this embodiment of this application, the divergence constraint of the N latent identity vectors may be further determined. In this case, the determining a loss of the target network model according to the identity-anonymized image may include: determining the loss of the target network model according to the identity-anonymized image and the divergence constraint.

[0095] For example, the divergence constraint L_{kl} of the N latent identity vectors may be determined by using the following formula (1):

$$L_{kl} = \frac{1}{2}\sum_{i \in N}(\mu_i^2 + \sigma_i^2 - \log \sigma_i^2 - 1) \qquad (1)$$

μ_i is a mean corresponding to the i^{th} latent identity vector in the N latent identity vectors, and σ_i is a variance corresponding to the i^{th} latent identity vector in the N latent identity vectors.

[0096] The formula (1) is only an example, and a manner of determining the divergence constraint of the N latent identity vectors in this embodiment of this application includes but is not limited to the formula (1), and for example, may be another manner of calculating a divergence constraint, for example, by using a variant of the formula (1).

[0097] In this embodiment of this application, the divergence constraint L_{kl} is applied to the N latent identity vectors. In this way, after training, the projection module fully learns identity information, and the first space of the projection module meets a standard Gaussian distribution. Therefore, during subsequent anonymization, the first space may be directly sampled to generate N latent identity vectors conforming to a standard Gaussian distribution, for generating a

virtual identity vector.

[0098] In some embodiments, the second space is obtained by performing nonlinear mapping on the first space, and is a complex non-Gaussian distribution. As shown in FIG. 5, after identity information is mapped to the first space, it is found that a distribution of an intermediate latent space, namely, the second space *W*, is not uniform in this case, and real identity vectors gather to a plurality of different centers and do not overlap generated virtual identity vectors. Therefore, no reasonable face identity can be generated based on the virtual identity vectors. Therefore, this embodiment of this application proposes use of a contrast loss to constrain a latent vector of the second space, namely, a *W* space (namely, the first virtual identity vector), so that latent vectors from the same identity gather and repel latent vectors from different identities, and all latent vectors are evenly distributed in the entire space.

[0099] Based on this, in this embodiment of this application, an identity loss may be further determined in the following manner:

Step 1: Obtain a third training image.

Step 2: Process the third training image by using a projection reference module to obtain N second virtual identity vectors.

Step 3: Determine the identity loss according to the N first virtual identity vectors and the N second virtual identity vectors.

[0100] The third training image and first training image are two different images of a first target. For example, the third training image and the first training image are two different face images of the same user.

[0101] The projection reference module and the projection module have the same network structure, and are updated according to the projection module. For example, the projection reference module is updated according to momentum of the projection module, that is, the projection reference module is updated slowly with an update of the projection module.

[0102] For example, the projection reference module may be updated according to the following formula (2):

$$P\theta'(t)=(1-\Delta)*P\theta'(t-1)+\Delta*P\theta(t) \qquad (2)$$

[0103] $P\theta'(t)$ is a projection reference module parameter obtained after the t^{th} update, $P\theta'(t-1)$ is a projection reference module parameter obtained after the $(t-1)^{th}$ update, $P\theta(t)$ is a projection module parameter obtained after the t^{th} update, and Δ is a small value, for example, 0.01.

[0104] As shown in FIG. 9, in the model training process, to determine the identity loss, in this embodiment of this application, the projection reference module with a network structure completely the same as that of the projection module is set, to constrain the first virtual identity vector outputted by the projection module. For example, the first training image is inputted to the projection module to obtain the N first virtual identity vectors of the first training image, and the third training image is inputted to the projection reference module to obtain the N second virtual identity vectors of the third training image. The first training image and the third training image are images of the same target, and the projection module and the projection reference module have the same network structure. In this way, after model training is completed, a difference between the N first virtual identity vectors corresponding to the first training image and the N second virtual identity vectors is small. Based on this, the projection module in the target network model may be trained according to the N second virtual identity vectors and the N first virtual identity vectors corresponding to the first training image, so that the projection module can generate a virtual identity vector that meets a requirement.

[0105] Manners of determining the identity loss according to the N first virtual identity vectors and the N second virtual identity vectors in step 1 include but are not limited to the following manners:

Manner 1: Determine differences between the N first virtual identity vectors and the N second virtual identity vectors at different resolutions, and determine a sum of the differences or an average value of the differences as the identity loss. For example, N is 3, a difference 1 between a first virtual identity vector 1 and a second virtual identity vector 1 is determined, a difference 2 between a first virtual identity vector 2 and a second virtual identity vector 2 is determined, and a difference 3 between a first virtual identity vector 3 and a second virtual identity vector 3 is determined. A sum of the difference 1, the difference 2, and the difference 3 is determined as the identity loss, or an average value of the difference 1, the difference 2, and the difference 3 is determined as the identity loss.

Manner 2: In this embodiment of this application, N dynamic lists K are designed, and the dynamic list stores representations of all different target identities (for example, face identities) in an entire training set in the second

space: a *W* + space. In this case, the identity loss may be determined in the following manner according to the N first virtual identity vectors and the N second virtual identity vectors:

[0106] Step 31: For the i^{th} first virtual identity vector in the N first virtual identity vectors, update, by using the i^{th} second virtual identity vector, a virtual identity vector corresponding to the first target in the i^{th} dynamic list.

[0107] The i^{th} dynamic list includes virtual identity vectors of different targets at the i^{th} resolution, and i is a positive integer ranging from 1 to N.

[0108] In this embodiment of this application, each of the N second virtual identity vectors corresponds to a dynamic list. For example, N is 3, and corresponds to a low resolution, a medium resolution, and a high resolution. In this case, there are also three dynamic lists: a first dynamic list corresponding to the low resolution, a second dynamic list corresponding to the medium resolution, and a third dynamic list corresponding to the high resolution.

[0109] Assuming that i is equal to 1, a virtual identity vector corresponding to the first target in the first dynamic list is updated by using the 1^{st} second virtual identity vector.

[0110] Assuming that i is equal to 2, a virtual identity vector corresponding to the first target in the second dynamic list is updated by using the 2^{nd} second virtual identity vector.

[0111] Assuming that i is equal to 3, a virtual identity vector corresponding to the first target in the third dynamic list is updated by using the 3rd second virtual identity vector.

[0112] Step 32: Determine, according to the i^{th} first virtual identity vector and the updated i^{th} dynamic list, an identity sub-loss corresponding to the i^{th} first virtual identity vector.

[0113] For example, as shown in FIG. 10, the first training image and the third training image are two different images of a first target j, the first training image Xj is inputted to the projection module to obtain N first virtual identity vectors Wj, and the third training image Xj' is inputted to the projection reference module to obtain N second virtual identity vectors Wj'. For the i^{th} resolution in N resolutions, the i^{th} dynamic list Ki includes second virtual identity vectors of different targets at the i^{th} resolution, and the i^{th} dynamic list Ki is updated in real time. For example, the i^{th} second virtual identity vector is used for updating a virtual identity vector kj corresponding to the first target j in the i^{th} dynamic list Ki, that is, kj is updated to Wj'. Then the identity sub-loss i corresponding to the i^{th} first virtual identity vector is determined according to the i^{th} second virtual identity vector and the updated i^{th} dynamic list.

[0114] A manner of determining the identity sub-loss corresponding to the i^{th} first virtual identity vector is not limited in this embodiment of this application.

[0115] For example, the identity sub-loss corresponding to the i^{th} first virtual identity vector is determined according to the i^{th} first virtual identity vector and the updated i^{th} dynamic list by using a loss method, for example, a center loss or a triplet loss.

[0116] In some embodiments, the determining an identity sub-loss corresponding to the i^{th} first virtual identity vector in the N first virtual identity vectors may include the following steps:

[0117] Step 321: Obtain a first ratio of the i^{th} second virtual identity vector to a first preset value, multiply the first ratio by the i^{th} first virtual identity vector to obtain a first result, and perform an exponential operation on the first result to obtain a first operation value.

[0118] Step 322: Obtain a second ratio of each second virtual identity vector to the first preset value in the updated i^{th} dynamic list, for each second ratio, multiply the second ratio by the corresponding i^{th} first virtual identity vector to obtain a second result, and perform an exponential operation on the second result to obtain a second operation value corresponding to each second virtual identity vector.

[0119] Step 323: Determine a sum of second operation values corresponding to all second virtual identity vectors, obtain a third ratio of the first operation value to the sum, and perform a logarithmic operation on the third ratio to obtain a third operation value.

[0120] Step 324: Determine a negative number of the third operation value as the identity sub-loss corresponding to the i^{th} first virtual identity vector.

[0121] For example, w^j is an anchor point, the j^{th} item in Ki is a positive sample, and others are negative samples. An identity sub-loss L_c is determined by using a contrast loss in an information noise contrastive noise (InfoNCE) form. InfoNCE is a loss function for modifying autoregression based on mutual information.

[0122] For example, the identity sub-loss $L_c(i)$ corresponding to the i^{th} first virtual identity vector is determined according to the following formula (3):

$$L_c(i) = -log\frac{exp(w^j \cdot \frac{K[j]}{\tau})}{\sum_{k=1}^{K} exp(w^k \cdot \frac{K[k]}{\tau})} \qquad (3)$$

w^j is the i^{th} first virtual identity vector of the first target j, $K[j]$ is the i^{th} second virtual identity vector of the first target j, τ is the first preset value, $K[k]$ is the i^{th} second virtual identity vector corresponding to the k^{th} target in the i^{th} dynamic list, w^k is a first virtual identity vector corresponding to the k^{th} target, and K is a total quantity of targets included in the i^{th} dynamic list.

[0123] Step 33: Determine, as the identity loss of the target network model, a sum of identity sub-losses respectively corresponding to the N first virtual identity vectors.

[0124] After the identity sub-loss corresponding to the i^{th} first virtual identity vector is determined according to step 32, the sum of the identity sub-losses respectively corresponding to the N first virtual identity vectors is determined as the identity loss. For example, N is 3. An identity sub-loss corresponding to each of the three first virtual identity vectors is determined according to the foregoing method, and then a sum of identity sub-losses corresponding to the three first virtual identity vectors is determined as the identity loss of the model.

[0125] In this embodiment of this application, after the identity loss in the model training process is determined according to the foregoing method, the determining the loss of the target network model according to the identity-anonymized image and the divergence constraint includes the following steps:

[0126] determining the loss of the target network model according to the identity-anonymized image, the divergence constraint, and the identity loss.

[0127] For example, a reconstruction loss between the identity-anonymized image and the second training image is determined, and the loss of the target network model is determined according to the reconstruction loss, the divergence constraint, and the identity loss.

[0128] In an example, a difference between the identity-anonymized image and the second training image is determined as the reconstruction loss. For example, a sum of differences between pixels of the identity-anonymized image and corresponding pixels of the second training image is determined as the reconstruction loss.

[0129] In another example, the reconstruction loss L_{rec} is determined according to the following formula (4):

$$L_{rec} = \left| Y_{s,t} - X_t \right|_1 \qquad (4)$$

$Y_{s,t}$ is the identity-anonymized image, X_t is the second training image, and $|\ |_1$ is a 1-norm operation.

[0130] After the reconstruction loss L_{rec} is determined according to the foregoing steps, the loss of the target network model is determined according to the reconstruction loss, the divergence constraint, and the identity loss. For example, a weighted sum of the reconstruction loss, the divergence constraint, and the identity loss is determined as a final loss of the target network model.

[0131] In some embodiments, to improve model training accuracy, this embodiment of this application further includes determining an identity contrast loss of the identity-anonymized image, for example, including the following steps:

[0132] Step A: Determine a first distance between the identity-anonymized image and the first training image, a second distance between the identity-anonymized image and the second training image, and a third distance between the first training image and the second training image.

[0133] Step B: Determine a contrast loss according to the first distance, the second distance, and the third distance.

[0134] The first distance, the second distance, and the third distance may be determined by using any distance method, for example, a cosine distance.

[0135] Example 1: After the first distance, the second distance, and the third distance are determined according to step A, a sum of the first distance, the second distance, and the third distance is determined as the contrast loss.

[0136] Example 2: A sum of the first distance and a square of a difference between the second distance and the third distance is determined, and a difference between a preset value and the sum is determined as the contrast loss.

[0137] In an example, the contrast loss L_{ICL} is determined according to the following formula (5):

$$L_{ICL} = 1 - \cos\left(z_{id}(Y_{s,t}), z_{id}(X_s)\right) + \left(\cos\left(z_{id}(Y_{s,t}), z_{id}(X_t)\right) - \cos\left(z_{id}(X_s), z_{id}(X_t)\right)\right)^2 \qquad (5)$$

z_{id} indicates a 512-dimensional identity vector representation, of an image X, that is extracted from a pre-trained face recognition model, $\cos(z_{id}(Y_{s,t}), z_{id}(X_s))$ is the first distance between the identity-anonymized image and the first training image, $\cos(z_{id}(Y_{st}), z_{id}(X_t))$ is the second distance between the identity-anonymized image and the second training image, and $\cos(z_{id}(X_s), z_{id}(X_t))$ is the third distance between the first training image and the second training image.

[0138] After the contrast loss L_{ICL} is determined according to the foregoing steps, the loss of the target network model

is determined according to the reconstruction loss, the divergence constraint, the identity loss, and the contrast loss. For example, a weighted sum of the reconstruction loss, the divergence constraint, the identity loss, and the contrast loss is determined as the loss of the target network model.

[0139] In some embodiments, in a case that the fusion module is an adversarial network, in the model training process, an adversarial loss of the model is further determined. For example, the adversarial loss is determined according to the identity-anonymized image and the first training image.

[0140] For example, the adversarial loss L_{GAN} is determined according to the following formula (6):

$$L_{GAN} = \min_{G} \max_{D} E\left[\log\left(D(X_s)\right)\right] + E\left[\log\left(1 - D\left(Y_{s,t}\right)\right)\right] \qquad (6)$$

[0141] D is a discriminator, G is a generator, E(*) indicates an expected value of a distribution function, $D(X_s)$ is a discrimination result of the discriminator on the first training image X_s, and $D(Y_{s,t})$ is a discrimination result of the discriminator on the identity-anonymized image $Y_{s,t}$.

[0142] After the adversarial loss L_{GAN} is determined according to the foregoing steps, the loss of the target network model may be determined according to the reconstruction loss, the divergence constraint, the identity loss, the contrast loss, and the adversarial loss. For example, a weighted sum of the reconstruction loss, the divergence constraint, the identity loss, the contrast loss, and the adversarial loss is determined as the loss of the target network model.

[0143] Weight values corresponding to the reconstruction loss, the divergence constraint, the identity loss, the contrast loss, and the adversarial loss are not limited in this embodiment of this application, and may be determined according to an actual requirement.

[0144] In some embodiments, a weighting operation is performed on the reconstruction loss, the divergence constraint, the identity loss, the contrast loss, and the adversarial loss according to the following formula (7) to obtain the loss L_{total} of the target network model:

$$L_{total} = L_{GAN} + 10 * L_{rec} + 5 * L_{ICL} + L_c + 0.0001 * L_{kl} \qquad (7)$$

[0145] A weight corresponding to each loss in the formula (7) is an example, and the weight corresponding to each loss in this embodiment of this application includes but is not limited to that shown in the formula (7), and may be determined as needed.

[0146] In some embodiments, to improve training accuracy of the target network model, a loss other than the losses described in the foregoing embodiment may be determined. This is not limited in this embodiment of this application, and may be determined according to an actual requirement.

[0147] It can be learned from the foregoing descriptions that, in this embodiment of this application, first virtual identity vectors corresponding to different resolutions are generated to implement identity anonymization. This can increase a resolution for anonymization. For example, an anonymization result at a resolution of 1024^2 may be generated. In addition, little picture artifact is produced, and fidelity is high. In addition, in this embodiment of this application, model training does not rely on a key regression model or a segmentation model, that is, a face area in an image is not removed, and a posture, details, and occlusion in an original picture are retained.

[0148] In this embodiment of this application, in the application process of the target network model, the first training image is projected to the target space by using the projection module to obtain the N first virtual identity vectors, so that the target network model can fully learn identity information in the image; attribute vector extraction is performed on the second training image to obtain the M attribute vectors, so that the target network model fully learns attribute information in the image; and image generation is performed based on the N first virtual identity vectors and the M attribute vectors to obtain the identity-anonymized image of the second training image. In this way, a trained model can generate an image carrying virtual identity information while ensuring that attribute information of an original image remains unchanged; That is, this application provides a new target network model. With the foregoing training method, the target network model can learn identity information in the first training image, so that the target network model can independently generate a virtual identity, and the target network model also fully learns attribute information in the second training image. In an entire learning process, a face area in the image does not need to be removed, and no real identity information is required for guidance either. In addition, the target network model is trained by using a specific supervised target in a face swapping task, thereby improving fidelity and a resolution of an identity-anonymized image generated by the target network model, so that the trained target network model can generate a high-quality identity-anonymized image.

[0149] The foregoing describes in detail the model training method in this application with reference to FIG. 3 to FIG.

10. The following describes in detail an identity anonymization method in this application with reference to FIG. 11 to FIG. 13.

[0150] FIG. 11 is a schematic flowchart of an identity anonymization method according to an embodiment of this application. The identity anonymization method shown in FIG. 11 is to perform identity anonymization by using the foregoing trained target network model. As shown in FIG. 11, the method includes:

S401: Perform sampling on a target space of a projection module in the target network model to obtain N virtual identity vectors, N being a positive integer.

[0151] It can be learned from the foregoing embodiment that, in this embodiment of this application, the projection module is trained by using a first training image, so that the projection module fully learns identity information in the first training image. In this way, in actual use, sampling may be performed on the target space of the projection module to obtain the N virtual identity vectors.

[0152] Implementations of S401 include but are not limited to the following:

Manner 1: Sampling is performed based on a mean and a variance of the target space of the trained projection module to obtain the N virtual identity vectors. For example, random sampling is performed on a variance of the target space, and then the variance is added to the mean of the target space to obtain a virtual identity vector. The foregoing step may be repeatedly performed to obtain the N virtual identity vectors.

Manner 2: The target space includes a first space and a second space, the target network model includes a second projection unit. In this case, sampling may be performed on the target space of the projection module in the target network model in the following manner to obtain the N virtual identity vectors:

performing sampling on the first space to obtain N latent identity vectors; and projecting, by using the second projection unit, the N latent identity vectors to the second space to obtain the N virtual identity vectors.

[0153] In this embodiment of this application, during actual anonymization, the first projection unit in the projection module is no longer used, and only the second projection unit in the projection module is used for projection. For example, as shown in FIG. 12, sampling is performed on a first space Z conforming to a standard Gaussian distribution, to obtain N latent identity vectors, and then the N latent identity vectors are inputted to the second projection unit. The second projection unit projects the N latent identity vectors to the W space to obtain the N virtual identity vectors. In FIG. 12, for example, N is 3, and the second projection unit is a mapping network. However, the projection module in this embodiment of this application is not limited to that shown in FIG. 12.

[0154] It can be learned from the foregoing descriptions that the first space is trained by using the first training image, so that a variance and a mean of the first space conform to a standard Gaussian distribution. In this case, sampling is first performed on the first space to generate the N latent identity vectors. For example, sampling is performed based on the mean and the variance of the first space to obtain the N latent identity vectors. Random sampling is performed on a variance of the first space, and then the variance is added to the mean of the first space to obtain a latent identity vector. The foregoing step may be repeatedly performed to obtain the N latent identity vectors. Then the N latent identity vectors are projected, by using the second projection unit, to the second space to obtain the N virtual identity vectors.

[0155] In some embodiments, the N virtual identity vectors respectively correspond to different resolutions. For example, N = 3, the 1^{st} virtual identity vector corresponds to a low resolution, the 2^{nd} virtual identity vector corresponds to a medium resolution, and the 3^{rd} virtual identity vector corresponds to a high resolution.

[0156] After the N virtual identity vectors are obtained according to the foregoing method, the following steps of S402 and S403 are performed to obtain an identity-anonymized image of a to-be-processed image.

[0157] S402: Perform, by using an attribute module in the target network model, attribute vector extraction on a to-be-processed image to obtain M attribute vectors, M being a positive integer.

[0158] In this embodiment of this application, the attribute module is configured to extract attribute information in the to-be-processed image.

[0159] In some embodiments, the attribute module includes an encoding unit and a decoding unit. In this case, attribute vector extraction may be performed on the to-be-processed image in the following manner to obtain the M attribute vectors:

[0160] inputting the to-be-processed image to the encoding unit to obtain feature information of the to-be-processed image; and inputting the feature information to the decoding unit to obtain the M attribute vectors.

[0161] In some embodiments, the encoding unit may include a plurality of feature extraction layers, and likewise, the decoding unit may also include a plurality of feature extraction layers. The feature extraction layer may include a convolutional layer or the like.

[0162] In some embodiments, a skip connection is implemented between at least one feature extraction layer in the encoding unit and at least one feature extraction layer in the decoding unit.

[0163] The generated M attribute vectors may correspond to different resolutions.

[0164] In some embodiments, the target network model is an autoencoder.

[0165] S403: Perform, by using a fusion module in the target network model, image generation based on the N virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the to-be-processed image.

[0166] According to the foregoing step, the generated N virtual identity vectors and M attribute vectors are inputted to the fusion module to obtain the identity-anonymized image of the to-be-processed image.

[0167] Implementations of S403 include but are not limited to the following examples:

[0168] Example 1: The N virtual identity vectors are spliced, the M attribute vectors are spliced, and a spliced virtual identity vector and a spliced attribute vector are combined and then inputted to the fusion module.

[0169] For example, the spliced virtual identity vector and the spliced attribute vector are concatenated and then inputted to the fusion module.

[0170] For another example, the spliced virtual identity vector and the spliced attribute vector are added up and then inputted to the fusion module.

[0171] Example 2: The fusion module includes a plurality of different resolution layers. In this case, according to resolutions corresponding to the N virtual identity vectors, the N virtual identity vectors may be inputted to a corresponding resolution layer as patterns, and the M attribute vectors may be inputted to a corresponding resolution layer as noises, to obtain the identity-anonymized image of the to-be-processed image.

[0172] In some embodiments, the fusion module is a StyleGAN2. In this case, as shown in FIG. 7, an AdaIN layer is included between two adjacent resolution layers of the fusion module. For example, affine transform is performed on a first virtual identity vector i+1; feature information outputted by the i^{th} resolution layer and an attribute vector i are combined and then inputted to the AdaIN layer along with a virtual identity vector i+1 obtained through affine transform; an AdaIN operation is performed, and an AdaIN operation result is inputted to the $(i+1)^{th}$ resolution layer.

[0173] The fusion module in this embodiment of this application may be alternatively an adversarial model, for example, a StyleGAN3 or a ProGAN. In some embodiments, the identity anonymization process in this embodiment of this application is described by using an example in which the second projection unit is a mapping network, the attribute module is an autoencoder, and the fusion module is a StyleGAN2.

[0174] For example, as shown in FIG. 13, sampling is performed on the first space Z of the projection module to obtain N latent identity vectors, for example, obtain three N latent identity vectors, the three N latent identity vectors respectively corresponding to three different resolutions: low, module, and high. Then the N latent identity vectors are inputted to the mapping network, and the mapping network projects the N latent identity vectors from the first space Z to the second space W to obtain N virtual identity vectors. In addition, a to-be-processed image Xt is inputted to the autoencoder, and the autoencoder processes the to-be-processed image Xt to generate M attribute vectors. Finally, the M attribute vectors are inputted to layers of the StyleGAN2 as noises, and the N virtual identity vectors are inputted to layers of the StyleGAN2 as patterns, to obtain an identity-anonymized image Ys,t, outputted by the StyleGAN2, of the to-be-processed image.

[0175] In the identity anonymization method provided in this embodiment of this application, sampling is performed on the target space of the projection module in the target network model to obtain the N virtual identity vectors; attribute vector extraction is performed, by using the attribute module in the target network model, on the to-be-processed image to obtain the M attribute vectors; and image generation is performed by using the fusion module in the target network model based on the N virtual identity vectors and the M attribute vectors to obtain the identity-anonymized image of the to-be-processed image. That is, the target network model in this embodiment of this application may generate an independent virtual identity, and during identity anonymization on the to-be-processed image, a face area in the to-be-processed image does not need to be removed, thereby improving fidelity of identity anonymization.

[0176] The foregoing describes in detail the method embodiments of this application with reference to FIG. 3 to FIG. 13. The following describes in detail apparatus embodiments of this application with reference to FIG. 14 and FIG. 15. FIG. 14 is a schematic block diagram of a model training apparatus according to an embodiment of this application. The training apparatus 10 may be a computing device, or a part of a computing device. As shown in FIG. 14, the model training apparatus 10 includes:

a projection unit 11, configured to project, by using a projection module in a target network model, a first training image to a target space to obtain N first virtual identity vectors, N being a positive integer;

an attribute unit 12, configured to perform, by using an attribute module in the target network model, attribute vector extraction on a second training image to obtain M attribute vectors, M being a positive integer;

a fusion unit 13, configured to perform, by using a fusion module in the target network model, image generation based on the N first virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the second training image; and

a training unit 14, configured to determine a loss of the target network model according to the identity-anonymized image, and train the target network model according to the loss.

[0177] In some embodiments, the projection module includes a first projection unit and a second projection unit, the target space includes a first space and a second space, and the projection unit 11 is further configured to: extract priori identity information of the first training image; project, by using the first projection unit, the priori identity information to the first space to obtain N latent identity vectors; and project, by using the second projection unit, the N latent identity vectors to the second space to obtain the N first virtual identity vectors.

[0178] In some embodiments, the projection unit 11 is further configured to: project, by using the first projection unit, the priori identity information as a mean and a variance of the first space; and perform sampling based on the mean and the variance of the first space to obtain the N latent identity vectors.

[0179] In some embodiments, the training unit 14 is further configured to: determine a divergence constraint of the N latent identity vectors; and determine the loss of the target network model according to the identity-anonymized image and the divergence constraint.

[0180] In some embodiments, the N first virtual identity vectors respectively correspond to different resolutions.

[0181] In some embodiments, the first projection unit is a variational autoencoder.

[0182] In some embodiments, the training unit 14 is further configured to: obtain a third training image, the third training image and the first training image being two different images of a first target; project, by using a projection reference module in the target network model, the third training image to the target space to obtain N second virtual identity vectors, the projection reference module and the projection module having the same network structure, and being updated according to the projection module; determine an identity loss according to the N first virtual identity vectors and the N second virtual identity vectors; and determine the loss of the target network model according to the identity-anonymized image, the divergence constraint, and the identity loss.

[0183] In some embodiments, the training unit 14 is further configured to: for the i^{th} second virtual identity vector in the N second virtual identity vectors, update, by using the i^{th} second virtual identity vector, a virtual identity vector corresponding to the first target in the i^{th} dynamic list, the i^{th} dynamic list including virtual identity vectors of different targets at the i^{th} resolution, and i being a positive integer ranging from 1 to N; determine, according to the i^{th} first virtual identity vector and the updated i^{th} dynamic list, an identity sub-loss corresponding to the i^{th} first virtual identity vector; and determine, as the identity loss, a sum of identity sub-losses respectively corresponding to the N first virtual identity vectors.

[0184] In some embodiments, the training unit 14 is further configured to: obtain a first ratio of the i^{th} second virtual identity vector to a first preset value, multiply the first ratio by the i^{th} first virtual identity vector to obtain a first result, and perform an exponential operation on the first result to obtain a first operation value; obtain a second ratio of each second virtual identity vector to the first preset value in the updated i^{th} dynamic list, for each second ratio, multiply the second ratio by the corresponding i^{th} first virtual identity vector to obtain a second result, and perform an exponential operation on the second result to obtain a second operation value corresponding to each second virtual identity vector; determine a sum of second operation values corresponding to all second virtual identity vectors, obtain a third ratio of the first operation value to the sum, and perform a logarithmic operation on the third ratio to obtain a third operation value; and determine a negative number of the third operation value as the identity sub-loss corresponding to the i^{th} first virtual identity vector.

[0185] In some embodiments, the attribute module includes an encoding unit and a decoding unit, and the attribute unit 12 is further configured to: perform, by using the encoding unit, feature extraction on the second training image to obtain feature information of the second training image; and decode, by using the decoding unit, the feature information to obtain M attribute vectors.

[0186] In some embodiments, a skip connection is implemented between at least one feature extraction layer in the encoding unit and at least one feature extraction layer in the decoding unit.

[0187] In some embodiments, the fusion module includes a plurality of different resolution layers, and the fusion unit 13 is further configured to: according to resolutions corresponding to the N first virtual identity vectors, input the N first virtual identity vectors to a corresponding resolution layer as patterns, and input the M attribute vectors to a corresponding resolution layer as noises, to obtain the identity-anonymized image of the second training image.

[0188] In some embodiments, the training unit 14 is further configured to: determine a reconstruction loss between the identity-anonymized image and the second training image; and determine the loss of the target network model based on the reconstruction loss, the divergence constraint, and the identity loss.

[0189] In some embodiments, the training unit 14 is further configured to: determine a first distance between the identity-anonymized image and the first training image, a second distance between the identity-anonymized image and the second training image, and a third distance between the first training image and the second training image; determine a contrast loss according to the first distance, the second distance, and the third distance; and determine the loss of the target network model based on the reconstruction loss, the divergence constraint, the identity loss, and the contrast loss.

[0190] In some embodiments, the training unit 14 is further configured to: determine a sum of the first distance and a square of a difference between the second distance and the third distance; and determine a difference between a preset value and the sum as the contrast loss.

[0191] In some embodiments, in a case that the fusion module is an adversarial network, the training unit 14 is further configured to: determine an adversarial loss according to the identity-anonymized image and the first training image; and determine a weighted sum of the reconstruction loss, the divergence constraint, the identity loss, the contrast loss, and the adversarial loss as the loss of the target network model.

[0192] It should be understood that the device embodiments and the method embodiments correspond to each other. For a similar description, refer to the method embodiments. To avoid repetition, details are not described herein again. For example, the apparatus shown in FIG. 14 may perform the embodiment of the model training method shown in FIG. 3, and the foregoing and other operations and/or functions of the modules in the apparatus are separately intended to implement the method embodiments corresponding to the computing device. For brevity, details are not described herein again.

[0193] FIG. 15 is a schematic block diagram of an identity anonymization apparatus according to an embodiment of this application. The identity anonymization apparatus 20 may be a computing device, or a part of a computing device. As shown in FIG. 15, the identity anonymization apparatus 20 includes:

[0194] a sampling unit 21, configured to perform sampling on a target space of a projection module in a target network model to obtain N virtual identity vectors, N being a positive integer;

[0195] an attribute unit 22, configured to perform, by using an attribute module in the target network model, attribute vector extraction on a to-be-processed image to obtain M attribute vectors, M being a positive integer; and

[0196] an anonymization unit 23, configured to perform, by using a fusion module in the target network model, image generation based on the N virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the to-be-processed image.

[0197] In some embodiments, the target space includes a first space and a second space, the target network model includes a second projection unit, and the sampling unit 21 is further configured to: perform sampling on the first space to obtain N latent identity vectors; and project, by using the second projection unit, the N latent identity vectors to the second space to obtain the N virtual identity vectors.

[0198] In some embodiments, a mean and a variance of the first space meet a standard Gaussian distribution, and the sampling unit 21 is further configured to perform sampling based on the mean and the variance of the first space to obtain the N latent identity vectors.

[0199] In some embodiments, the N virtual identity vectors respectively correspond to different resolutions.

[0200] In some embodiments, the attribute module includes an encoding unit and a decoding unit, and the attribute unit 22 is further configured to: perform, by using the encoding unit, feature extraction on the to-be-processed image to obtain feature information of the to-be-processed image; and decode, by using the decoding unit, the feature information to obtain M attribute vectors.

[0201] In some embodiments, a skip connection is implemented between at least one feature extraction layer in the encoding unit and at least one feature extraction layer in the decoding unit.

[0202] In some embodiments, the fusion module includes a plurality of different resolution layers, and the anonymization unit 23 is further configured to: according to resolutions corresponding to the N virtual identity vectors, input the N virtual identity vectors to a corresponding resolution layer as patterns, and input the M attribute vectors to a corresponding resolution layer as noises, to obtain the identity-anonymized image of the to-be-processed image.

[0203] It should be understood that the device embodiments and the method embodiments correspond to each other. For a similar description, refer to the method embodiments. To avoid repetition, details are not described herein again. For example, the apparatus shown in FIG. 15 may perform the embodiment of the identity anonymization method shown in FIG. 11, and the foregoing and other operations and/or functions of the modules in the apparatus are separately intended to implement the method embodiments corresponding to the computing device. For brevity, details are not described herein again.

[0204] The foregoing describes the apparatuses in the embodiments of this application from a perspective of functional modules with reference to the accompanying drawings. The functional modules may be implemented in a form of hardware, or may be implemented by instructions in a form of software, or may be implemented by a combination of hardware and software modules. For example, the steps of the method embodiments in the embodiments of this application may be performed by an integrated logic circuit in a processor and/or instructions in a form of software, and the steps of the methods disclosed with reference to the embodiments of this application may be directly performed by a hardware decoding processor, or may be performed by a combination of hardware and software modules in a decoding processor. Optionally, the software module may be located in a mature storage medium in the art, such as a random access memory, a flash memory, a read-only memory, a programmable read-only memory, an electrically-erasable programmable memory, and a register. The storage medium is located in the memory. The processor reads information in the memory and completes the steps of the foregoing method embodiments in combination with hardware thereof.

[0205] FIG. 16 is a schematic block diagram of a computing device according to an embodiment of this application. The computing device is configured to perform the foregoing method embodiments. As shown in FIG. 16, the computing device 30 may include:

[0206] a memory 31 and a processor 32, the memory 31 being configured to store a computer program 33 and transmit the program code 33 to the processor 32. In other words, the processor 32 may invoke the computer program 33 from the memory 31 and run the computer program 33, to implement the methods in the embodiments of this application.

[0207] For example, the processor 32 may be configured to perform the foregoing method steps according to instructions in the computer program 33.

[0208] In some embodiments of this application, the processor 32 may include but is not limited to:

[0209] a general processor, a digital signal processor (Digital Signal Processing, DSP), an application-specific integrated circuit (Application Specific Integrated Circuit, ASIC), a field programmable gate array (Field Programmable Gate Array, FPGA), another programmable logic device, a discrete gate, a transistor logic device, or a discrete hardware component.

[0210] In some embodiments, the storage module 31 includes, but is not limited to:

a non-volatile and/or volatile memory. The non-volatile memory may be a read-only memory (ROM), a programmable read-only memory (PROM), an erasable programmable read-only memory (EPROM), an electrically erasable programmable read-only memory (EEPROM) or a flash memory. The volatile memory may be a random access memory (RAM), used as an external cache. Through example but not limitative description, many forms of RAMs may be used, for example, a static random access memory (static RAM, SRAM), a dynamic random access memory (dynamic RAM, DRAM), a synchronous dynamic random access memory (synchronous DRAM, SDRAM), a double data rate synchronous dynamic random access memory (double data rate SDRAM, DDR SDRAM), an enhanced synchronous dynamic random access memory (enhanced SDRAM, ESDRAM), a synchronous link dynamic random access memory (synchlink DRAM, SLDRAM), and a direct rambus dynamic random access memory (direct rambus RAM, DR RAM).

[0211] In some embodiments of this application, the computer program 33 may be divided into one or more modules, and the one or more modules are stored in the memory 31 and executed by the processor 32 to perform the page recording method provided in this application. The one or more modules may be a series of computer program instruction segments capable of performing specific functions, and the instruction segments are used for describing execution processes of the computer program 33 on the computing device.

[0212] As shown in FIG. 16, the computing device 30 may further include:

a transceiver 34, where the transceiver 34 may be connected to the processor 32 or the memory 31.

[0213] The processor 32 may control the transceiver 34 to communicate with another device. For example, information or data may be sent to another device or receive information or data sent by another device. The transceiver 34 may include a transmitter and a receiver. The transceiver 34 may further include an antenna, and a quantity of the antenna can be one or more.

[0214] Various components of the computer device 30 are connected to each other by using a bus system. In addition to including a data bus, the bus system further includes a power bus, a control bus, and a status signal bus.

[0215] An embodiment of this application provides a computer storage medium, where the computer storage medium stores a computer program, and when the computer program is executed by a computer, the method in any one of the foregoing embodiments is implemented by the computer, or, a computer program product including instructions is further provided in the embodiments of this application. When the instructions run on a computer, the computer is caused to perform the method provided in the foregoing method embodiments.

[0216] According to an aspect of the embodiments of this application, a computer program product or a computer program is provided, the computer program product or the computer program including computer instructions, the computer instructions being stored in a computer-readable storage medium. A processor of a computer device reads the computer instructions from the computer-readable storage medium. The processor executes the computer instructions, so that the computer device performs the method in the foregoing various method embodiments.

[0217] The foregoing descriptions are merely a specific implementation of this application, but are not intended to limit the protection scope of this application. Any variation or replacement readily figured out by a person skilled in the art within the technical scope disclosed in this application shall fall within the protection scope of this application. Therefore, the protection scope of this application shall be subject to the protection scope of the claims.

Claims

1. A model training method, the method being performed by a computing device, and comprising:

projecting, by using a projection module in a target network model, a first training image to a target space to obtain N first virtual identity vectors, N being a positive integer;

performing, by using an attribute module in the target network model, attribute vector extraction on a second training image to obtain M attribute vectors, M being a positive integer;

performing, by using a fusion module in the target network model, image generation based on the N first virtual

identity vectors and the M attribute vectors to obtain an identity-anonymized image of the second training image; and

determining a loss of the target network model according to the identity-anonymized image, and training the target network model according to the loss.

2. The method according to claim 1, wherein the projection module comprises a first projection unit and a second projection unit, the target space comprises a first space and a second space, and the projecting, by using a projection module in a target network model, a first training image to a target space to obtain N first virtual identity vectors comprises:

extracting priori identity information of the first training image;

projecting, by using the first projection unit, the priori identity information to the first space to obtain N latent identity vectors; and

projecting, by using the second projection unit, the N latent identity vectors to the second space to obtain the N first virtual identity vectors.

3. The method according to claim 2, wherein the projecting, by using the first projection unit, the priori identity information to the first space to obtain N latent identity vectors comprises:

projecting, by using the first projection unit, the priori identity information as a mean and a variance of the first space; and

performing sampling based on the mean and the variance of the first space to obtain the N latent identity vectors.

4. The method according to claim 2 or 3, wherein the method further comprises:

determining a divergence constraint of the N latent identity vectors; and

the determining a loss of the target network model according to the identity-anonymized image comprises:

determining the loss of the target network model according to the identity-anonymized image and the divergence constraint.

5. The method according to claim 4, wherein the method further comprises:

obtaining a third training image, the third training image and the first training image being two different images of a first target;

projecting, by using a projection reference module in the target network model, the third training image to the target space to obtain N second virtual identity vectors, the projection reference module and the projection module having the same network structure, and being updated according to the projection module; and

determining an identity loss according to the N first virtual identity vectors and the N second virtual identity vectors; and

the determining the loss of the target network model according to the identity-anonymized image and the divergence constraint comprises:

determining the loss of the target network model according to the identity-anonymized image, the divergence constraint, and the identity loss.

6. The method according to claim 5, wherein the determining an identity loss according to the N first virtual identity vectors and the N second virtual identity vectors comprises:

for the i^{th} second virtual identity vector in the N second virtual identity vectors, updating, by using the i^{th} second virtual identity vector, a virtual identity vector corresponding to the first target in the i^{th} dynamic list, the i^{th} dynamic list comprising virtual identity vectors of different targets at the i^{th} resolution, and i being a positive integer ranging from 1 to N;

determining, according to the i^{th} first virtual identity vector and the updated i^{th} dynamic list, an identity sub-loss corresponding to the i^{th} first virtual identity vector; and

determining, as the identity loss, a sum of identity sub-losses respectively corresponding to the N first virtual identity vectors.

7. The method according to claim 6, wherein the determining, according to the i^{th} first virtual identity vector and the updated i^{th} dynamic list, an identity sub-loss corresponding to the i^{th} first virtual identity vector comprises:

obtaining a first ratio of the i^{th} second virtual identity vector to a first preset value, multiplying the first ratio by the i^{th} first virtual identity vector to obtain a first result, and performing an exponential operation on the first result to obtain a first operation value;

obtaining a second ratio of each second virtual identity vector to the first preset value in the updated i^{th} dynamic list, for each second ratio, multiplying the second ratio by the corresponding i^{th} first virtual identity vector to obtain a second result, and performing an exponential operation on the second result to obtain a second operation value corresponding to each second virtual identity vector;

determining a sum of second operation values corresponding to all second virtual identity vectors, obtaining a third ratio of the first operation value to the sum, and performing a logarithmic operation on the third ratio to obtain a third operation value; and

determining a negative number of the third operation value as the identity sub-loss corresponding to the i^{th} first virtual identity vector.

8. The method according to any one of claims 1 to 7, wherein the attribute module comprises an encoding unit and a decoding unit, and the performing, by using an attribute module in the target network model, attribute vector extraction on a second training image to obtain M attribute vectors comprises:

performing, by using the encoding unit, feature extraction on the second training image to obtain feature information of the second training image; and

decoding, by using the decoding unit, the feature information to obtain M attribute vectors.

9. The method according to any one of claims 1 to 7, wherein the fusion module comprises a plurality of different resolution layers, and the performing, by using a fusion module in the target network model, image generation based on the N first virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the second training image comprises:

according to resolutions corresponding to the N first virtual identity vectors, inputting the N first virtual identity vectors to a corresponding resolution layer as patterns, and inputting the M attribute vectors to a corresponding resolution layer as noises, to obtain the identity-anonymized image of the second training image.

10. The method according to claim 5, wherein the determining the loss of the target network model according to the identity-anonymized image, the divergence constraint, and the identity loss comprises:

determining a reconstruction loss between the identity-anonymized image and the second training image; and

determining the loss of the target network model based on the reconstruction loss, the divergence constraint, and the identity loss.

11. The method according to claim 10, wherein the method further comprises:

determining a first distance between the identity-anonymized image and the first training image, a second distance between the identity-anonymized image and the second training image, and a third distance between the first training image and the second training image; and

determining a contrast loss according to the first distance, the second distance, and the third distance; and

the determining the loss of the target network model based on the reconstruction loss, the divergence constraint, and the identity loss comprises:

determining the loss of the target network model based on the reconstruction loss, the divergence constraint, the identity loss, and the contrast loss.

12. The method according to claim 11, wherein the determining a contrast loss according to the first distance, the second distance, and the third distance comprises:

determining a sum of the first distance and a square of a difference between the second distance and the third distance; and

determining a difference between a preset value and the sum as the contrast loss.

13. The method according to claim 11, wherein in a case that the fusion module is an adversarial network, the determining the loss of the target network model based on the reconstruction loss, the divergence constraint, the identity loss, and the contrast loss comprises:

determining an adversarial loss according to the identity-anonymized image and the first training image; and
determining a weighted sum of the reconstruction loss, the divergence constraint, the identity loss, the contrast loss, and the adversarial loss as the loss of the target network model.

14. An identity anonymization method, the method being performed by a computing device, and comprising:

performing sampling on a target space of a projection module in a target network model to obtain N virtual identity vectors, N being a positive integer;
performing, by using an attribute module in the target network model, attribute vector extraction on a to-be-processed image to obtain M attribute vectors, M being a positive integer; and
performing, by using a fusion module in the target network model, image generation based on the N virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the to-be-processed image.

15. The method according to claim 14, wherein the target space comprises a first space and a second space, the target network model comprises a second projection unit, and the performing sampling on a target space of a projection module in a target network model to obtain N virtual identity vectors comprises:

performing sampling on the first space to obtain N latent identity vectors; and
projecting, by using the second projection unit, the N latent identity vectors to the second space to obtain the N virtual identity vectors.

16. The method according to claim 15, wherein a mean and a variance of the first space meet a standard Gaussian distribution, and the performing sampling on the first space to obtain N latent identity vectors comprises:
performing sampling based on the mean and the variance of the first space to obtain the N latent identity vectors.

17. A model training apparatus, the apparatus comprising:

a projection unit, configured to project, by using a projection module in a target network model, a first training image to a target space to obtain N first virtual identity vectors, N being a positive integer;
an attribute unit, configured to perform, by using an attribute module in the target network model, attribute vector extraction on a second training image to obtain M attribute vectors, M being a positive integer;
a fusion unit, configured to perform, by using a fusion module in the target network model, image generation based on the N virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the second training image; and
a training unit, configured to determine a loss of the target network model according to the identity-anonymized image, and train the target network model according to the loss.

18. An identity anonymization apparatus, the apparatus comprising:

a sampling unit, configured to perform sampling on a target space of a projection module in a target network model to obtain N virtual identity vectors, N being a positive integer;
an attribute unit, configured to perform, by using an attribute module in the target network model, attribute vector extraction on a to-be-processed image to obtain M attribute vectors, M being a positive integer; and
an anonymization unit, configured to perform, by using a fusion module in the target network model, image generation based on the N virtual identity vectors and the M attribute vectors to obtain an identity-anonymized image of the to-be-processed image.

19. A computer device, comprising a processor and a memory;

the memory being configured to store a computer program; and
the processor being configured to implement the method according to any of claims 1 to 13, or the method according to any of claims 14 to 16 when executing the computer program.

20. A computer-readable storage medium, storing a computer program, the computer program, when executed, causing a computer device to implement the method according to any one of claims 1 to 13, or the method according to any of claims 14 to 16.

21. A computer program product, comprising a computer program or instruction, the computer program or instruction,

when executed by a processor, implementing the method according to any of claims 14 to 16, or the method according to any of claims 1 to 13.

FIG. 1A

FIG. 1B

FIG. 1C

FIG. 1D

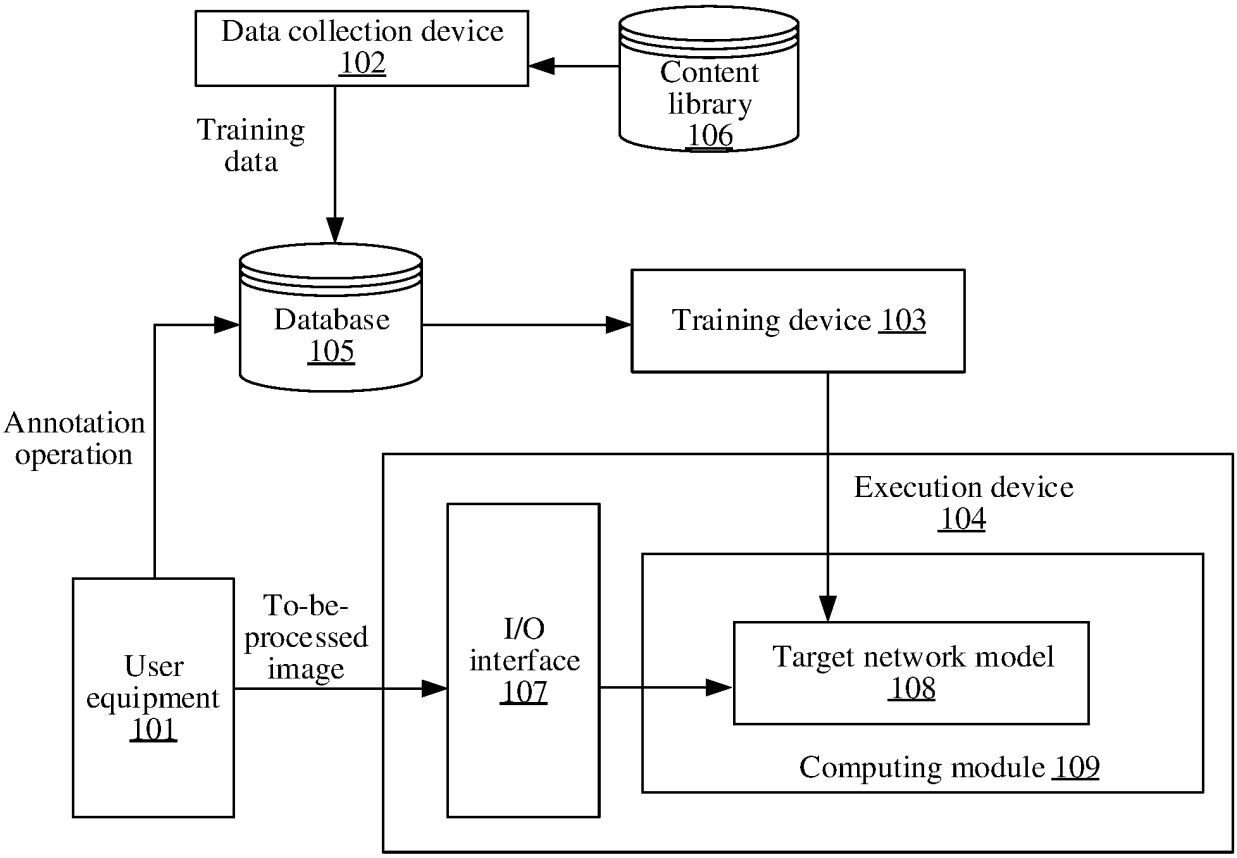
Data collection device 102
Content library 106
Training data
Database 105
Training device 103
Annotation operation
Execution device 104
User equipment 101
To-be-processed image
I/O interface 107
Target network model 108
Computing module 109

FIG. 2

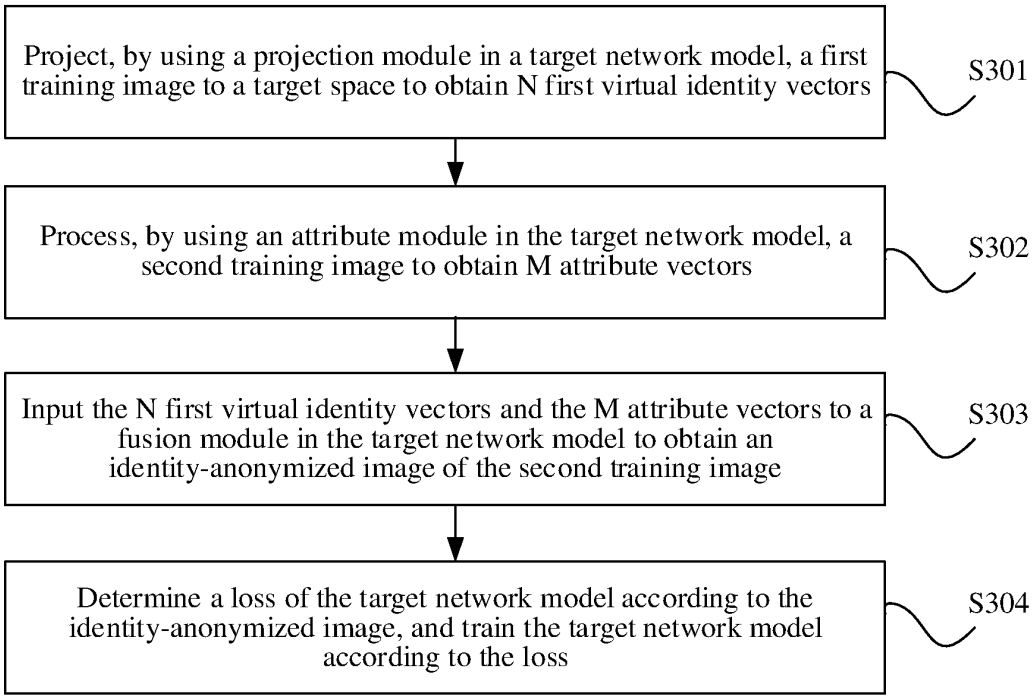

FIG. 3

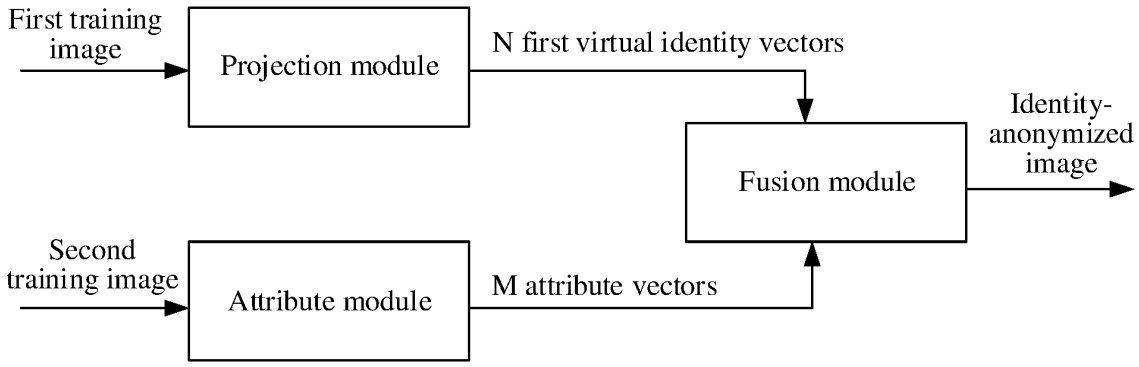

FIG. 4

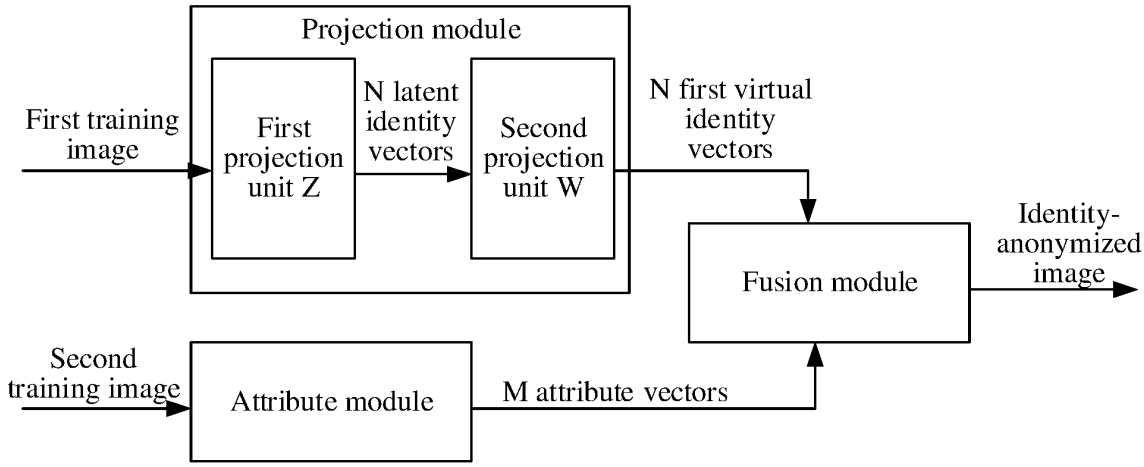

FIG. 5

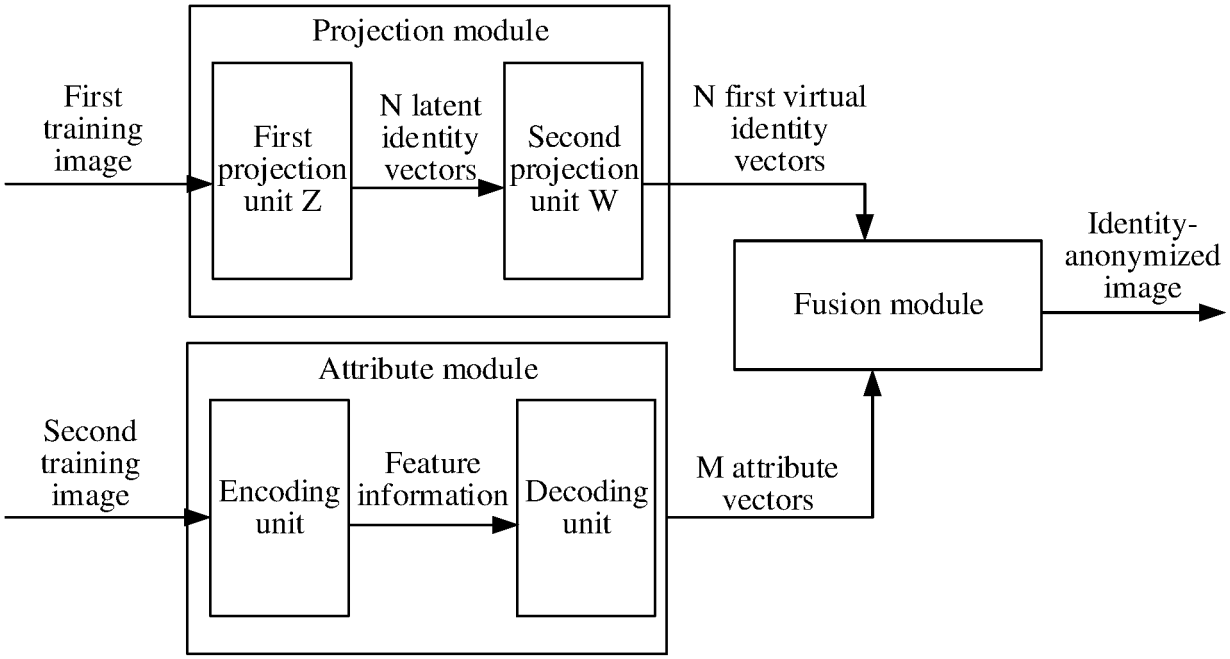
Projection module
First training image
First projection unit Z
N latent identity vectors
Second projection unit W
N first virtual identity vectors
Fusion module
Identity-anonymized image
Attribute module
Second training image
Encoding unit
Feature information
Decoding unit
M attribute vectors

FIG. 6

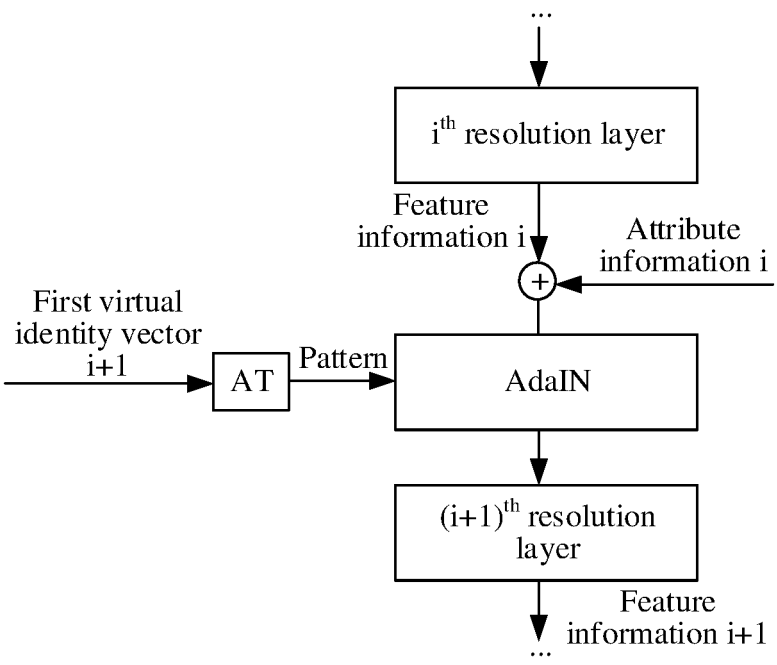
...
i^th resolution layer
Feature information i
Attribute information i
First virtual identity vector i+1
AT
Pattern
AdaIN
(i+1)^th resolution layer
Feature information i+1
...

FIG. 7

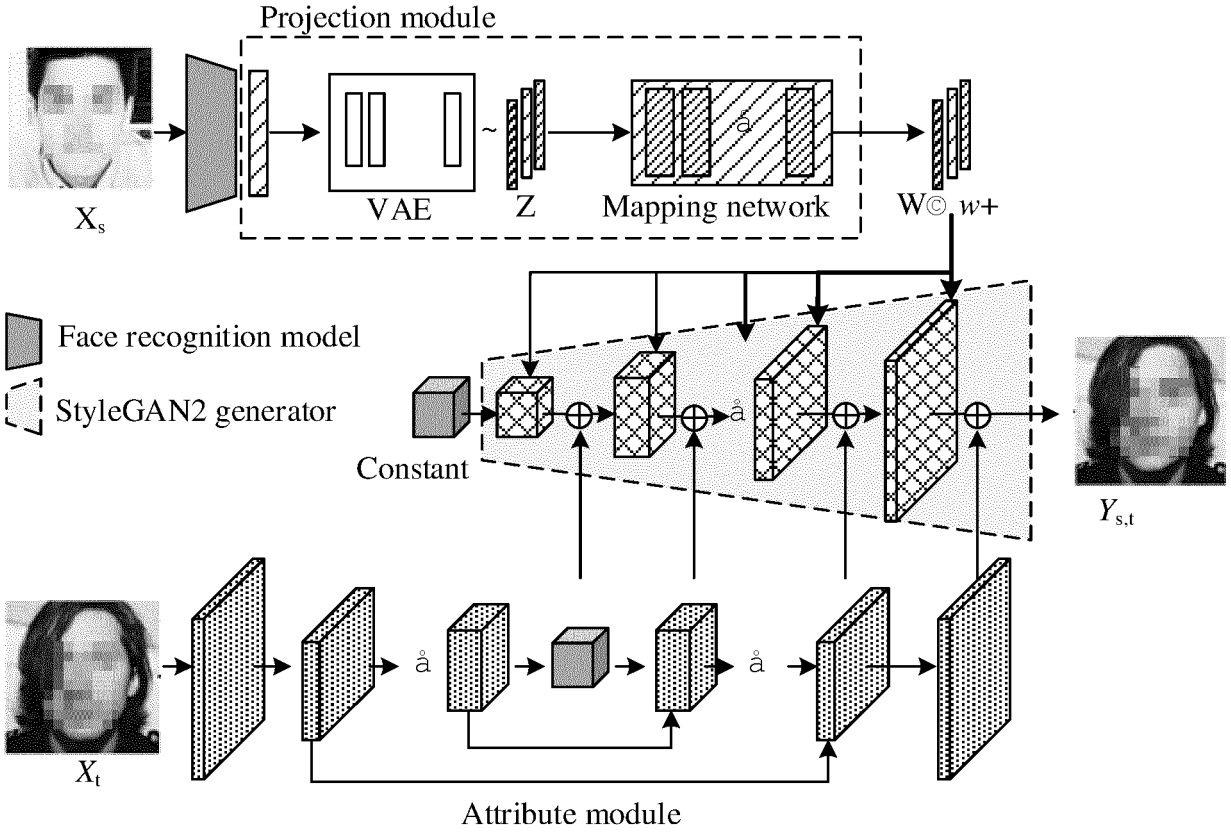
Projection module
X_s
VAE
Z
Mapping network
W© w+
Face recognition model
StyleGAN2 generator
Constant
$Y_{s,t}$
X_t
Attribute module

FIG. 8

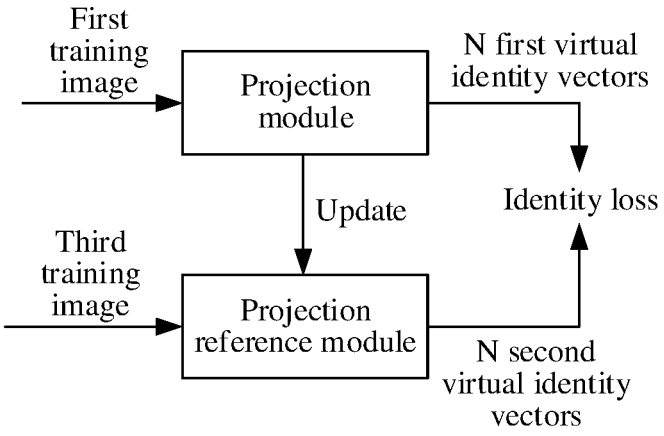
First training image
Projection module
N first virtual identity vectors
Update
Identity loss
Third training image
Projection reference module
N second virtual identity vectors

FIG. 9

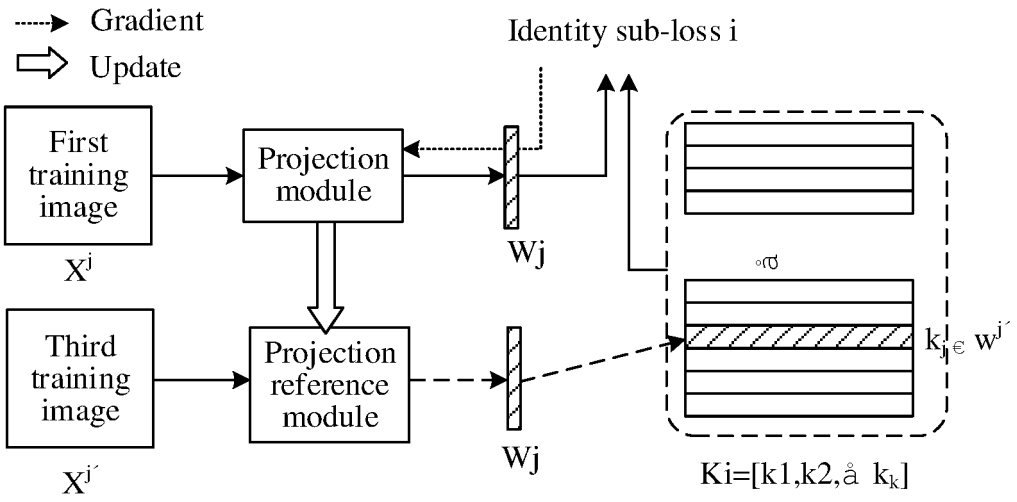

FIG. 10

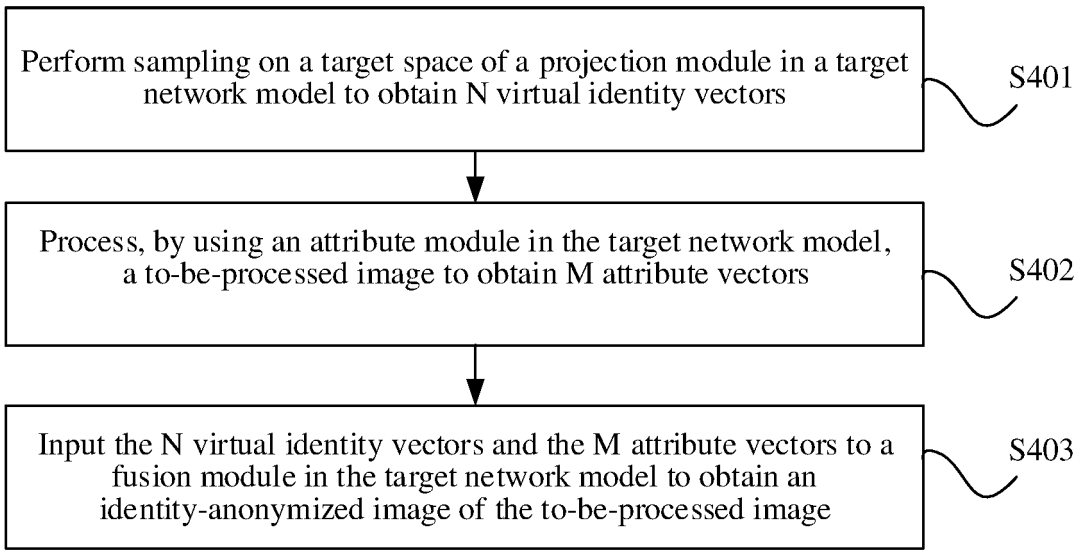

FIG. 11

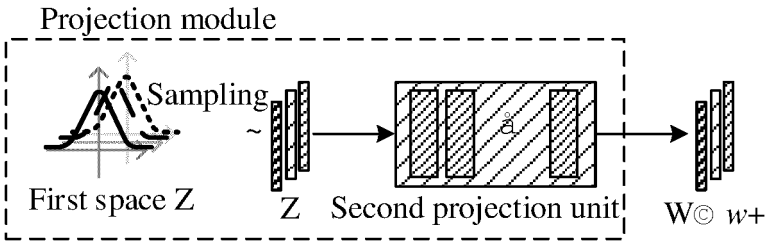

FIG. 12

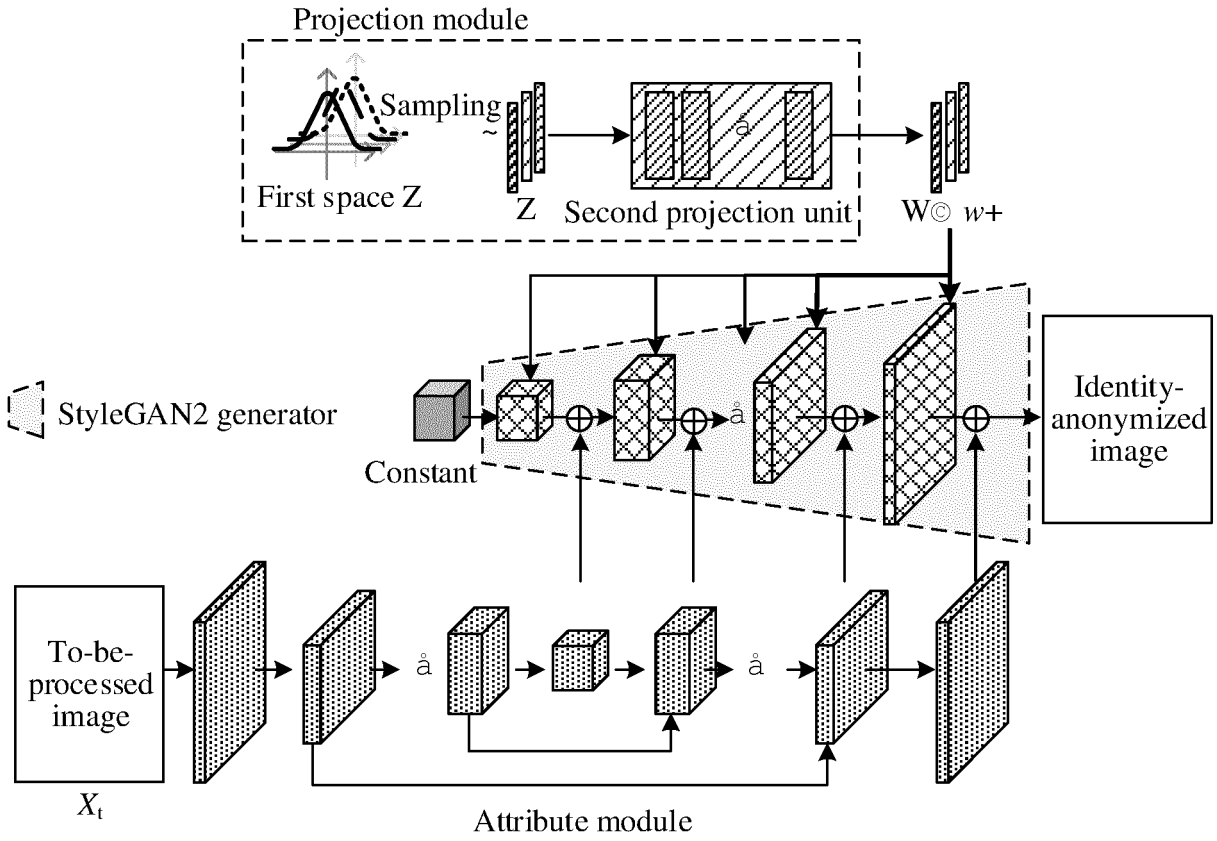
Projection module
Sampling
First space Z
Z
Second projection unit
W© w+
StyleGAN2 generator
Constant
Identity-anonymized image
To-be-processed image
X_t
Attribute module

FIG. 13

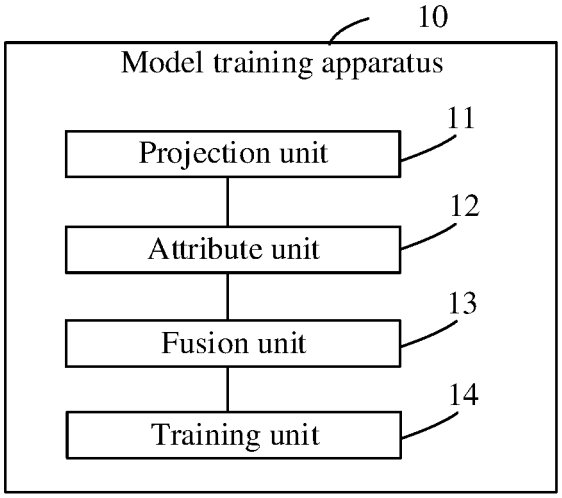
10
Model training apparatus
11
Projection unit
12
Attribute unit
13
Fusion unit
14
Training unit

FIG. 14

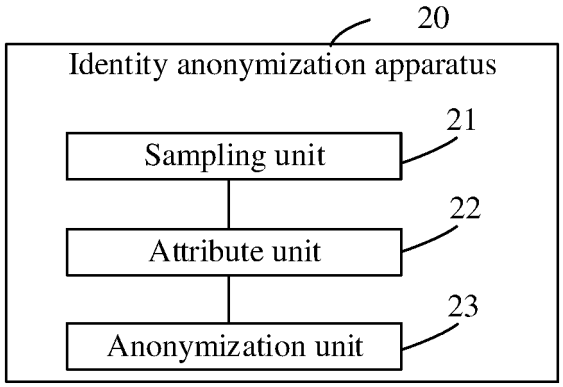
20
Identity anonymization apparatus
21
Sampling unit
22
Attribute unit
23
Anonymization unit

FIG. 15

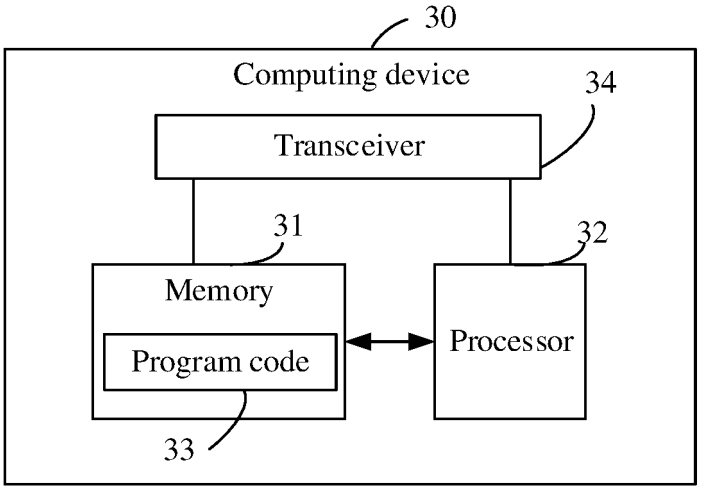
30
Computing device
34
Transceiver
31
32
Memory
Program code
Processor
33

FIG. 16

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2022/111704

A. CLASSIFICATION OF SUBJECT MATTER

G06F 21/62(2013.01)i; G06T 3/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F; G06T; G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI, IEEE: 身份, 脸, 面部, 匿名, 隐藏, 隐私, 换脸, 交换, 替换, 虚拟, 特征, 向量, 属性, 背景, 投影, 空间, 先验, 隐向量, 融合, 合并, 模型, 训练, 损失, 约束, 重构, 散度, 参考, 对抗网络, De-Identification, identity, face, GAN, VAE, CVAE, anonymous, hidden, privacy, change, virtual, feature, vector, attribute, background, space, fusion, merge, model, training, loss, constraint, reconstruct, diversity, reference

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 113033511 A (INSTITUTE OF AUTOMATION, CHINESE ACADEMY OF SCIENCES) 25 June 2021 (2021-06-25) description, paragraphs [0019]-[0031]	1-21
Y	CN 114120041 A (JINAN UNIVERSITY et al.) 01 March 2022 (2022-03-01) description, paragraph [0012]	1-21
Y	WO 2021258920 A1 (BAIGUOYUAN TECHNOLOGY (SINGAPORE) PTE LTD.) 30 December 2021 (2021-12-30) description, pages 4-16	1-21
Y	CN 113642409 A (SHANGHAI JIAO TONG UNIVERSITY) 12 November 2021 (2021-11-12) description, paragraphs [0048]-[0160]	1-21
A	CN 114139198 A (HANGZHOU DIANZI UNIVERSITY) 04 March 2022 (2022-03-04) entire document	1-21

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 November 2022	**07 December 2022**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2022/111704

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 113033511 A	25 June 2021	None	
CN 114120041 A	01 March 2022	None	
WO 2021258920 A1	30 December 2021	CN 111783603 A	16 October 2020
CN 113642409 A	12 November 2021	None	
CN 114139198 A	04 March 2022	None	

Form PCT/ISA/210 (patent family annex) (January 2015)

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- CN 202210234385 **[0001]**